Identification and Characterization of Novel Drug Resistance Loci in Plasmodium falciparum

Professor Dyann F. Wirth Daria Van Tyne

Identification and characterization of novel drug resistance loci in *Plasmodium falciparum*

ABSTRACT

Malaria has plagued mankind for millennia. Antimalarial drug use over the last century has generated highly drug-resistant parasites, which amplify the burden of this disease and pose a serious obstacle to control efforts. This dissertation is motivated by the simple fact that malaria parasites have become resistant to nearly every antimalarial drug that has ever been used, yet the precise genetic mechanisms of parasite drug resistance remain largely unknown. Our work pairs genomics-age technologies with molecular biology, genetics and molecular epidemiology in order to identify and characterize novel genes that contribute to drug resistance in *P. falciparum*.

In the Introduction, we highlight relevant opportunities and challenges in trying to identify and understand the genetic basis of malaria drug resistance as it emerges to currently used therapies. In Chapter One, we demonstrate how genome-wide association studies (GWAS) can be applied to *P. falciparum* in order to identify novel drug resistance loci. Functional follow-up revealed that overexpression of the novel candidate gene *PF10_0355* made parasites more resistant to the drugs halofantrine, lumefantrine and mefloquine. These findings show that *PF10_0355* plays a role in parasite drug response, as well as provides validation of our GWAS approach.

In Chapter Two, we further characterize *PF10_0355* and show that modulation of the

gene by either knockout or allelic replacement changes parasite drug sensitivity. Furthermore, we show that moderate changes in drug response measured in the short-term can have dramatic effects when parasites are competed with one another under drug pressure. In Chapter Three, we use an overexpression approach to functionally follow up other novel drug resistance genes generated by GWAS in *P. falciparum*. We find that overexpression is a useful way to begin to screen candidate drug resistance loci in the malaria parasite.

In Chapter Four, we use a DAPI-based *ex vivo* drug assay to monitor drug resistance among parasites circulating in Thiés, Senegal. In the future, we will look for genetic markers of parasite drug resistance in this population by GWAS. Finally, in the Discussion we present an essay about malaria evolution and eradication written for non-specialists. Our hope is that the work presented in this dissertation furthers understanding of drug resistance in the malaria parasite, both within and beyond the malaria research community.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This dissertation project required input and expertise from so many different people that I sometimes think of them all as a "malaria army" of sorts. The general of that army would have to be my thesis adviser Dyann Wirth, who opened doors, let me follow my interests, and gave me the guidance and support that I needed to be successful. I am forever grateful for all she taught me about being a thoughtful scientist, working as part of a team and managing so many different projects effectively and efficiently.

The other members of the Wirth Lab at HSPH were invaluable to my thesis work, and to the maintenance of my sanity throughout this project. I was fortunate to overlap with other graduate students who helped me find my way around the lab in the early days: Amanda Lukens, Julia Nunes, Meg Rush, Jenny Sims, Carolyn Dong, and Vishal Patel. They have been replaced with a newer group of students who have infected me with their enthusiasm: Jon Herman, Allison Demas, Leila Ross, and Rachel Daniels. I have also been lucky to work with senior scientists and post-docs who allowed me to learn immensely from their vast and varied experience: Sarah Volkman, Danny Milner, Johanna Daily, Clarissa Valim, Ulf Ribacke, Pamela Magistrado, and Angana Mukherjee. Without the thousands of hours of work by numerous technicians in the lab, many of my projects would not have been possible. Thank you to Kayla Barnes, Dave Rosen, Michele LeRoux, Nira Mahesh, Justin Becker, Meaghan Galligan, Tom Burke, Kate Fernandez, and Gilberto Ramirez.

I am very grateful to our next-door neighbors in the Duraisingh Lab, who were willing to exchange lab supplies for critical advice and help with experimental design: Jeff Dvorin, Brad Coleman, Saurabh Patel, and Amy Bei. Also, thank you to my graduate student counterparts in the Sabeti Lab at Harvard, Elaine Angelino and Danny Park, who developed the computational

tools to conduct genome-wide association studies in malaria, despite my constantly nagging them for data.

One of the best things about my thesis work was that it afforded me the opportunity to interact with so many amazing and brilliant scientists, at HSPH as well as the Broad Institute. Roger Wiegand, Daniel Neafsey, Steve Schaffner, Kevin Galinsky, and Lauren Young were invaluable resources and cheerful collaborators. I also had the opportunity to conduct fieldwork for my dissertation in Senegal, which I found immensely rewarding and unforgettable. I am grateful to Professor Souleymane Mboup, Daouda Ndiaye, Ousmane Sarr, Ambroise Ahoudi, and Theo Sène for welcoming me into their laboratory and showing me what "field research" is like.

I would like to thank the members of my Dissertation Advisory Committee for their input and support throughout my thesis research: Eric Rubin, Manoj Duraisingh, Pardis Sabeti and Dan Hartl. I am also thankful to the members of my Dissertation Exam Committee, for participating in my defense: Eric Rubin, Matt Marti, Bill Hanage and Mike Ferdig.

I am grateful for the ample administrative support at HSPH that allowed me to focus entirely on research projects: Andi Sabaroff, Carmen Mejia, Shannon Joyce, Maura Meagher, Rebeckah Orton, Ramona Gonski and Colleen Saville. I am also grateful for support from the former and current BPH Program Directors, Michael Grusby and Marianne Wessling-Resnick.

My fellow BPH graduate students have made sure I had fun during what little time I spent outside of the lab: Prerna Bhargava, Ediz Calay, Celene Chang, Khanichi Charles, Chris Ford, Regina Joice, Sihao Liu, and Karell Pelle. And a special thanks to my BPH "partner in crime," Franzi Graf. I was also fortunate to participate in the Leder Human Biology Program,

and am thankful to Connie Cepko for her mentorship and insight at various points throughout my graduate career.

I worked in a variety of research labs before arriving at Harvard, and I am enormously appreciative of all the advice and guidance, not to mention all the recommendation letters, that I received from my previous mentors: Eric Eberhardt, Uli Göringer, Robert Batey, and Miguel Navarro.

Partway through my graduate career I was welcomed into a Harvard undergraduate community as a resident tutor, and it has been a truly amazing experience. I am lucky to have been able to interact with all of the extraordinary people I have met at Lowell House, especially Diana Eck, Dorothy Austin, Brett Flehinger, and Elizabeth Terry. I am also grateful to two non-biologist friends, Frank Centinello and KC Forcier, who helped to keep me grounded throughout my time in graduate school and reminded me that there is a world beyond malaria research and a world outside of Harvard.

I never would have thought that I could earn a PhD, much less at a place like Harvard, were it not for the support and encouragement of my family. My parents, Chester and Natalie, and my sisters, Mary, Kasey and Fiona, have always been enthusiastic cheerleaders and I am grateful for their love. My in-laws John and Donna Bruno, my sisters-in-law Emily and Sarah, and my brothers-in-law Joe and Mike, have all embraced me and welcomed me into their families with open arms, and I am fortunate for their support as well.

Finally, I am grateful to my husband Jonathan, who has made me both a better scientist and a better person. Thank you for your love and encouragement, for your constant words of wisdom, and for making me laugh. I am truly blessed.

DEDICATION

This dissertation is dedicated to my parents, Chester and Natalie Van Tyne. Thank you for all of your love and support.

INTRODUCTION

Watching Evolution in Real-Time: Identifying and understanding the genetic basis of malaria drug resistance as it emerges

Note: This Introduction was composed by Daria Van Tyne

Summary

Malaria is a significant burden on global health, and the evolution of drug resistant parasites greatly impedes efforts to treat and control this disease. The malaria community is uniquely positioned to watch evolution in real-time, as parasites with reduced sensitivity to artemisinin combination therapies (ACTs) have arisen but full-blown, large-scale, and widespread resistance has not yet been detected. Current research efforts are aimed at identifying mutations that cause ACT resistance, in order to inform proper drug use and provide biomarkers for resistance monitoring. This review discusses opportunities and challenges in trying to identify and understand the genetic basis of antimalarial drug resistance as it emerges.

The Road to Resistance: Definitions and Assumptions

Malaria in humans can be caused by infection with any one of five different *Plasmodium* species; here we will focus on infection with *Plasmodium falciparum*, because it causes the most malaria-associated deaths worldwide, predominantly among African children under five years old [1]. Because *P. falciparum* infection is often acute and symptomatic, it is also most often treated with antimalarial drugs. Widespread treatment with many different antimalarials since the 1950s has generated resistance to nearly every drug that has been used (Figure 1). Current first-line treatment for *P. falciparum* infection includes a derivative of artemisinin and a longer-lasting partner drug [2]. Unfortunately, parasites showing delayed clearance after ACT treatment have recently been observed [3], and it appears to be only a matter of time before full-blown ACT resistance develops and spreads throughout the world.

For the purposes of this introduction, *drug* will be defined as any molecule that hinders or stops the growth of *P. falciparum* parasites, either by directly killing the parasite (in the case of cytocidal drug effects) or by inhibiting parasite growth (cytostatic effects) [4]. *Drug resistance* is the ability of parasites to sustain growth and persist despite the presence of a drug. *Fitness* is the extent to which a parasite can proliferate either in the presence or absence of antimalarial drugs. *Selective pressure* is the restriction on parasite growth imposed by some external force, such as antimalarial drug use. This pressure is called selective because it *selects* a subset of parasites that can withstand the imposed force and eliminates all other parasites from the population.

It should be noted that *resistance* and *fitness* are both used here as relative terms. Malaria parasites can be “more resistant” or “more fit” than other parasites if they are able to sustain growth in an environment containing a higher dose of drug, or are able to grow faster than other

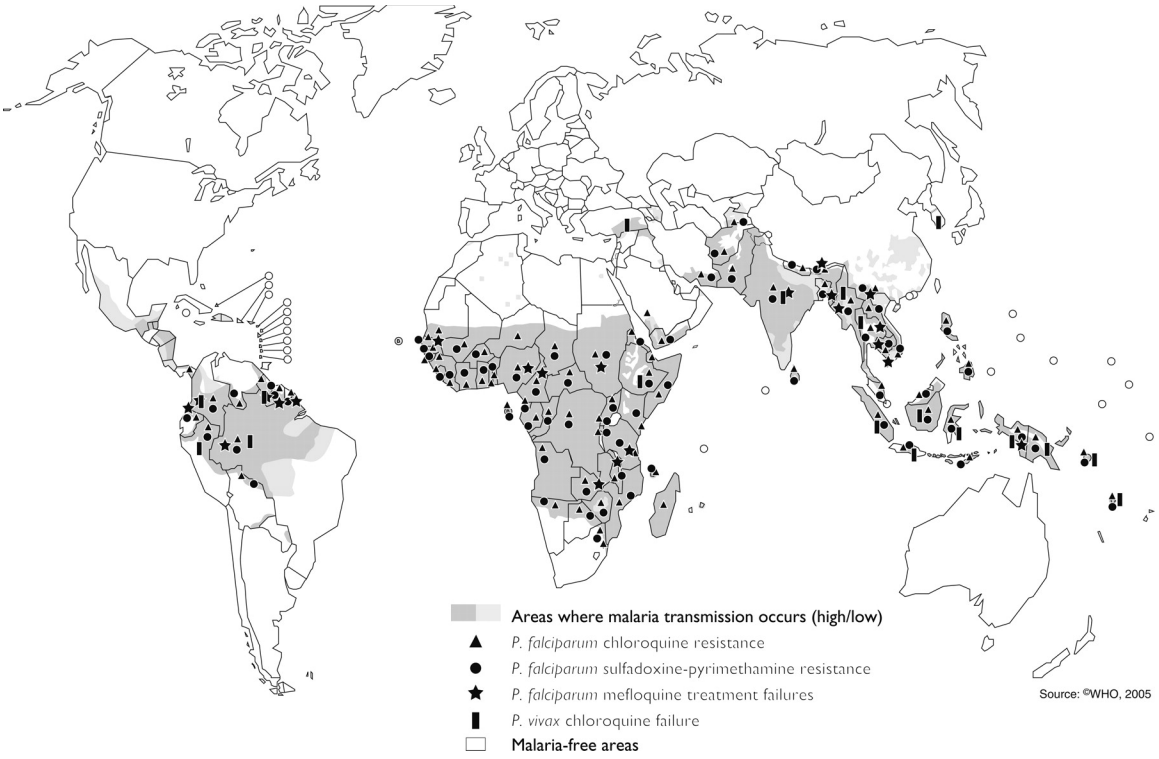


Figure 1. Malaria transmission, drug resistance and treatment failure in 2005. Figure reprinted with permission from [5].

parasites within a particular host environment. The World Health Organization defines clinical malaria drug resistance as the persistence of parasites within a patient after appropriate administration and absorption of a drug [6]. This is a useful *in vivo* definition; however, parasite drug susceptibility can also be measured *in vitro* or *ex vivo* using a number of standard assays [7] [8] [9], and parasites can be classified as "sensitive" or "resistant" according to these tests. Additionally, parasite fitness and selective pressures depend heavily upon the context in which they are considered. For example, fitness can refer to growth within a human, growth within a mosquito, or the ability of parasites to be transmitted from one host to another. Selective pressures abound throughout the malaria life cycle, as the parasite must exist in dynamic and highly selective host environments (Figure 2), and naturally imposed bottlenecks restrict the transition of parasites from one life cycle stage to the next [10].

For the purposes of this introduction, we will assume that there is a detectable genetic basis for resistance to most, if not all, antimalarial drugs. Non-mutational mechanisms of drug resistance exist in other systems [11], and may also exist in malaria [12]; however, these may be more difficult to elucidate than drug resistance that is driven by genetic variants, such as point mutations (also called single nucleotide polymorphisms or SNPs) or copy number variations (CNVs). In malaria, there are several well-known examples of both mutations and copy number variation that underlie antimalarial drug resistance. Mutations in drug targets like dihydrofolate reductase (*dhfr*), or in the drug resistance transporters *pfcrt* and *pfmdr1*, are understood to cause resistance to pyrimethamine, chloroquine and mefloquine, respectively [14] [15]. In the case of pyrimethamine and chloroquine, resistance mutations within *dhfr* and *pfcrt* appear to have arisen in a small number of independent geographic sites and spread globally. This is known because resistance mutations can be mapped back to their likely origins by looking for evidence of

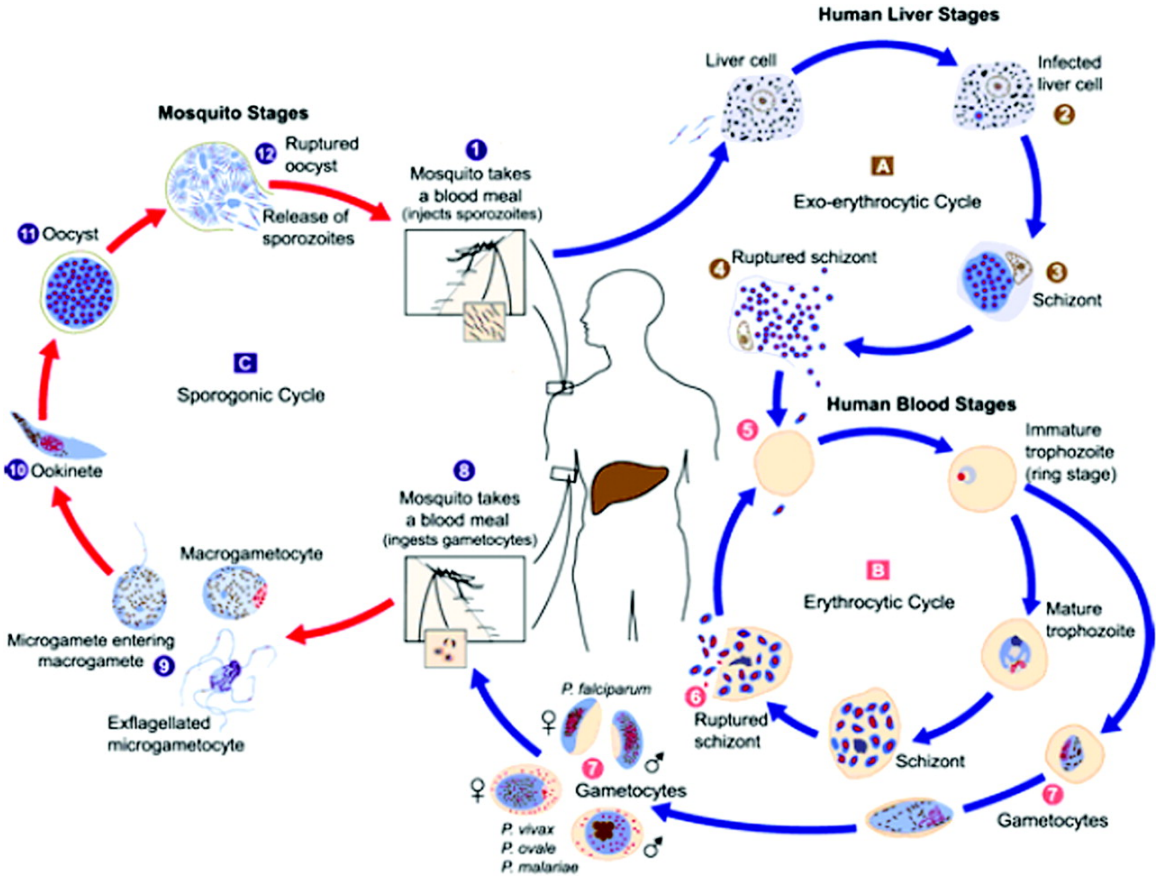


Figure 2. *Plasmodium* development within human and mosquito hosts. Reprinted with permission from [13].

reduced recombination, often called a *signature of selection*, which surrounds these genetic loci in the parasite's genome [16].

Finally, we will assume that mutations are occurring in the malaria genome at a near-constant rate. Estimates of precise mutation and recombination rates in *P. falciparum* vary widely, but we will focus on the idea that the parasite is constantly acquiring mutations and then testing whether each mutation is "contextually beneficial," meaning whether it improves parasite fitness in the current environment, as compared to other parasites without the mutation. Nearly all mutations cause no benefit to parasite fitness or are deleterious, and are therefore eliminated from the population. Occasionally, a mutation occurs that makes parasites better able to withstand a selective pressure, such as treatment with a particular antimalarial drug. At the end of the day, timing is everything: mutations must occur in the right place and at the right time in order to improve parasite fitness and thus remain in the population.

Road Signs: Identifying malaria drug resistance genes in the post-genome era

The genome sequence of *P. falciparum* was completed and published ten years ago, beginning the "genomics era" of malaria research [17]. Genomics-age technologies like microarrays and whole-genome sequencing have been widely used over the past decade to study many different aspects of malaria parasite biology, including drug resistance [18] [19]. The overall goal of these efforts is to better understand the basic biology underlying parasite drug response, as well as to develop better surveillance and monitoring tools in order to combat resistance in clinical settings. As resistance appears to be emerging to ACTs that are currently used worldwide, researchers are well positioned to harness the power of genomics in order to watch *P. falciparum* evolve in real time. Various approaches that researchers have used to search

for the genetic basis of antimalarial drug resistance are summarized below, along with discussion of the most relevant features of each approach to the current challenge of identifying ACT resistance mutations as they are emerging.

One tool to look for the genetic basis of a trait such as drug resistance is to use linkage analysis to examine trait inheritance among the progeny of a genetic cross. Parents that have differing phenotypes of interest are crossed, and the phenotypes of their progeny can be mapped back to parental loci using quantitative trait locus (QTL) mapping. Linkage analysis relies on a "genetic map" of markers (such as SNPs or microsatellites) that vary between the two parents, and these analyses are most effective when they employ a map with maximum marker density and test as many distinct progeny as possible. Three genetic crosses have been conducted in *P. falciparum* to date [20], one of which was used to identify a locus associated with chloroquine resistance [21]. This locus was originally mapped to a 400-kilobase stretch of chromosome 7, and it took nearly a decade to identify the *pfcrt* gene within this region as the underlying cause of the QTL signal and the principal driver of chloroquine resistance [22].

Because linkage studies employ two genetically distinct parents, every parasite studied will possess only one of two possible alleles of every gene in their genome. In a case where drug resistance is caused by a single mutation or polymorphism, crossing a clearly sensitive parasite line with a clearly resistant line will yield progeny that are either sensitive or resistant, with no intermediate phenotypes present. But if resistance is caused by a combination of polymorphisms at multiple loci, the progeny would likely display intermediate drug responses depending on which mutations they inherit, and numerous loci containing many different genes may be identified. This appears to be the case for artemisinin response measured in the parents and progeny of the HB3 x Dd2 parasite cross [23]. As reduced response to ACTs is currently

emerging but high-level resistance has not yet arisen, it may be difficult to apply linkage analysis to identify the genetic basis of ACT resistance. Other challenges to this approach include the lack of a strong *in vitro* correlate of delayed parasite clearance, as well as the technical and cost limitations of conducting the genetic cross. Linkage-based studies are generally a useful tool to identify the genetic basis of numerous parasite phenotypes, but this approach may not be effective for understanding ACT resistance until these challenges can be overcome.

Genome-wide association studies (GWASs) are complementary to linkage-based studies for understanding the genetic basis of malaria phenotypes such as drug resistance. GWASs use the natural diversity present within parasite populations to identify mutations that are associated with a phenotype of interest. Similar to linkage studies, GWASs work best when there are large phenotypic differences among the parasites studied, when many parasites are analyzed, and when the "genetic map" used is as dense as possible. GWASs in *P. falciparum* to date have employed SNP arrays [24] [25] [26], and more recently, whole-genome sequencing [27], in order to identify mutations that are associated with drug response among *in vitro* culture-adapted parasite lines. Low linkage disequilibrium, particularly among African isolates of *P. falciparum* [28], has two consequences for GWASs: first, SNP arrays can only test mutations that are present on the array and generally cannot assay other mutations that are not present; and second, association signals often map to small genomic regions that contain one or a few genes, rather than large regions containing many genes. GWASs in malaria have successfully identified known antimalarial resistance loci using as few as 45 parasite isolates [27], and have also identified numerous novel loci associated with drug resistance. Functional follow-up has been carried out for one recently identified novel locus associated with parasite response to halofantrine [25].

Similar follow-up will be required to validate other novel hits and to distinguish between true hits and false positive associations.

The GWAS approach is made more powerful when it is combined with tests that scan the genome for signatures of recent selection, and leverage this information to identify loci that are both under positive selection and also associated with the phenotype of interest [29]. These include long-range haplotype (LRH) tests such as integrated haplotype score (iHS) [30], and cross-population extended haplotype heterozygosity (XP-EHH) [31]. Selection-based tests look for areas of the genome that have not yet been broken down by mutation or recombination, but it is important to note that these tests alone cannot identify the source of selection. However, because antimalarial drug use is a strong selective pressure that has been recently imposed, drug use is a useful working hypothesis for the force underlying signals of recent selection found within the malaria genome. Recent studies have paired tests for selection with GWASs to identify loci that appear to be under recent positive selection only in drug resistant *P. falciparum* isolates [27], as well as genomic regions that are under selection and also associated with prolonged ACT clearance rates [32]. The latter study identified a 35-kilobase locus on chromosome 13 that contains over a dozen candidate genes and is associated with delayed parasite clearance in Thailand and Cambodia. It is possible that one or more of the genes within this locus cause delayed parasite clearance, but one can only speculate until functional validation experiments are carried out.

One final approach for understanding the genetic basis of ACT resistance is resistance selection, whereby a population of parasites is exposed to drug pressure *in vitro* and resistant parasite lines are generated and analyzed [33]. Resistance selection has been paired with whole-genome sequencing [34] or genomic tiling arrays [35] to identify the genetic basis of resistance

to numerous compounds. While this is a powerful approach, it should be noted that both genetic and transcriptional differences likely exist between parasites circulating in malaria-infected patients and culture-adapted lines growing *in vitro* [36] [37]. Furthermore, *in vitro* protocols may not accurately mirror the dynamics of resistance selection when it occurs in the wild. Selection for resistance to artemisinin and its derivatives in *P. falciparum* has been tried many times and in many different ways [38] [39] [40], and these studies have generated numerous candidate genes. Unfortunately, mutations in these candidate genes do not seem to correlate well with clinical observations of delayed parasite clearance [41], suggesting that their effect may be limited to culture-adapted parasites and the *in vitro* laboratory setting.

All three studies mentioned above—linkage analysis, association studies, and *in vitro* resistance selection—are useful for generating hypotheses and identifying candidate resistance genes, as they all take an unbiased approach and consider the entire genome when looking for markers of parasite drug resistance. There could be many potential roads to ACT resistance in malaria, and it seems fitting that as many relevant approaches as possible are employed to try and understand the genetic basis of resistance before it becomes widespread.

Forks in the Road: Understanding malaria drug resistance as it emerges

The malaria life cycle is complex, requiring the parasite to take up residence in numerous different environments within both human and mosquito hosts, and undergoing multiple population bottlenecks in order to pass from one developmental stage to the next. And yet somehow despite this complexity, and despite treatment and control efforts imposed over the last century, malaria parasites remain a serious burden on human health. Malaria certainly is a formidable foe, and its adaptability is clearly on display when considered in the context of the evolution of drug resistance. This section presents various considerations that may aid in the

discussion of how malaria drug resistance evolves. Knowledge of how drug resistance emerges in other diseases, as well as the biological consequences of being drug resistant, could better inform our understanding of how parasite resistance is emerging to ACTs.

It is important to first recognize the differences between antimalarial drug use during the last century and throughout the "natural history" of malaria evolution. Over the last century, synthetic compounds have been developed and deployed one at a time to try and kill all the parasites that cause a malaria infection. This imposes strong selective pressure on a diverse population of parasites within a patient, and favors the rapid development of high-level resistance [42]. Somewhat paradoxically, the emergence of drug resistant *P. falciparum* parasites seems to have accelerated throughout the 20th century, just as drug targets have become better defined and more specific [43]. Throughout malaria's natural history, parasites were likely exposed to many different naturally-occurring compounds within the human bloodstream and mosquito mid-gut that would exert antimalarial effects, similar to the case of naturally-occurring antibiotics among soil microbial communities [44]. The effects of these naturally derived compounds on inhibiting parasite growth may have been small, but nonetheless parasites had to evolve resistance or tolerance to them in order to persist. Additionally, *Artemisia* plant extracts that contain artemisinin (and likely many other compounds) have been used to treat fever in China for hundreds of years [45]. It is only in the last half-century that artemisinin has been isolated as the compound within these plants that is most effective at killing malaria parasites, leading to its isolation and combination with other synthetic compounds to constitute ACTs. It seems that drug exposure is not a new phenomenon, but the way that parasites now encounter drugs is far removed from their natural and historical contexts.

It is well known that copy number variation (CNV) plays a major role in modulating many different biological processes in malaria, including drug resistance [46]. One way that parasites are able to adapt to drug pressure is by rapidly increasing the copy number and thereby increasing the expression of genes within the pathway or pathways targeted by a particular drug. This appears to be the case for *GTP cyclohydrolase I* amplification in *P. falciparum*, which drives the parasite folate biosynthesis pathway and appears to be a mechanism for resistance to the antifolate drugs pyrimethamine and sulfadoxine [47]. Another example can be found in amplification of the *pfmdr1* locus, which increases parasite resistance to mefloquine, quinine and artemisinin, presumably by promoting drug efflux [48] [49]. Additionally, point mutations have been discovered within *pfmdr1* that seem to make parasites more resistant to chloroquine and amodiaquine [50] [51]. The interplay between CNVs and point mutations has been explored in other systems, such as host adaptation in double-stranded DNA viruses and antibiotic resistance in bacteria. Poxviruses appear to employ "gene accordions" to rapidly increase the copy number of a target gene and then accrue different mutations within each copy [52]. Once a highly beneficial mutation occurs in one of the amplified genes, the additional copies are lost and the accordion collapses back down. A recent study of antibiotic resistance in *Salmonella typhimurium* found that gene amplification was also followed by selection for a resistance-conferring mutation and subsequent de-amplification [53]. It seems possible that CNVs in *P. falciparum* similarly facilitate adaptive evolution, and that observations of amplifications and mutations within important malaria drug resistance genes represent different time points along this evolutionary trajectory.

Studies of drug resistance mutations in bacterial and viral systems have suggested that they often make organisms less fit [54] [55], though a possible exception to this rule in cancer

has been described [56]. The precise way in which resistance mutations impact parasite fitness can be measured by numerous methods, but one relevant measure is the effectiveness of parasite transmission to mosquitoes through conversion to sexual stages (called gametocytes). The impact of resistance mutations on the transmission potential of parasites has been studied in various rodent malaria models, where it was shown that resistant parasites are effectively transmitted following drug treatment, but can be outcompeted by sensitive parasites in the absence of drug pressure [57] [58]. Similar effects have been observed in *P. falciparum*, where patients harboring drug-resistant infections were found to have increased gametocyte levels following treatment with chloroquine and sulfaxoxine/pyrimethamine [59] [60] [61]. Other studies of *P. falciparum* populations across many parasite generations have suggested that in the absence of drug pressure, many drug resistance mutations incur a fitness cost over time [62] [63] [64] [65]. Although drug resistance mutations may make parasites less fit when they compete with sensitive lines in the absence of drug pressure, this effect clearly disappears when sensitive parasites are eliminated by drug treatment.

It also seems clear that many mutations that cause drug resistance are either a consequence of or necessitate the acquisition of compensatory mutations that increase parasite fitness, either in the presence or absence of drug pressure. In *P. falciparum*, mutations making parasites more tolerant to low-level concentrations of pyrimethamine might have allowed for less common and possibly less fit resistance mutations to arise within *dhfr* [66]. Parasites that show delayed clearance in patients treated with ACTs have also been described as "tolerant," suggesting that they could represent an intermediate stop along the road to full-blown ACT resistance [3] [67]. Conversely, selection for a mutation conferring high-level resistance could generate parasites that are initially unfit, but compensatory mutations can arise which quickly

restore fitness. Selection experiments in *P. chabaudi* have shown that a parasite line selected to be pyrimethamine resistant was initially unfit, but additional mutations led to increased fitness and allowed subsequent parasites to out-compete wild-type lines [68]. A separate study using transgenic yeast expressing the *P. falciparum* version of *dhfr* showed that the three most likely pathways to pyrimethamine resistance all involved an initial decrease in fitness, which correlated with large increases in drug resistance [69]. Finally, in addition to compensatory mutations occurring within the same drug target, fitness-restoring mutations can also arise in different genes within the same biological pathway, as in the case of *dhfr* mutations that were found in parasites with amplified *GTP cyclohydrolase I* [70]. In this case, it was suggested that increased copy number could compensate for decreased efficiency of the mutated dhfr enzyme by up-regulating the entire folate metabolism pathway. Whether compensatory mutations play a role in drug resistance acquisition or facilitate parasite survival after an un-fit resistance mutation occurs, they are clearly important and should be considered within the broader context of parasite adaptation to antimalarial drugs.

Other ways that *P. falciparum* can compensate for drug resistance are less defined, and highlight potential new areas of malaria drug resistance research efforts. Amplification of the chromosome 5 locus containing *pfmdr1*, in addition to making parasites drug resistant, also appears to alter the transcription of numerous genes, possibly through the action of neighboring regulatory factors [71]. These transcriptional changes could represent parasite adaptation to an altered, drug-resistant state. A separate study of parasites with different *pfct* mutations generated in an isogenic background found both expression and copy number changes in growth, development, signal transduction and transporter genes, which suggests numerous compensatory mechanisms for being chloroquine resistant [72]. A recent sequence-based GWAS for drug

resistance determinants in *P. falciparum* identified a number of genes that were under recent positive selection in pyrimethamine-resistant parasites, including chaperones and other proteins thought to be involved in lipid metabolism and ubiquitination [27]. These genes may represent novel biological pathways, such as stress response or protein turnover, that can help parasites survive under drug pressure and/or compensate for resistance mutations within *dhfr*. Lastly, parasites *in vitro* selected to be resistant to dihydroartemisinin showed the up-regulation of genes involved in antioxidant defense, as well as many chaperones [40]. Overall it seems that antimalarial drug resistance may involve many more genes and pathways than the subset which have been focused on thus far.

Finally, an important facet of malaria drug resistance that cannot be studied using culture-adapted parasite lines is the interplay between drug resistance and host immunity. The ability to clear chloroquine resistant *P. falciparum* has been associated with patient age [73], and acquired immunity seemed to improve the clearance of drug-resistant parasites in a study of Tanzanian children [74]. Three human polymorphisms were recently associated with enhanced ability to clear drug-resistant *P. falciparum* infections; two of these mutations were within pro-inflammatory and anti-inflammatory response genes [75]. One possible explanation for this phenomenon is that selection for drug-resistant parasites could cause reduced variation in the parasite's antigen repertoire, and as a result, resistant parasites would be less fit in the face of immune pressure [76]. While it is tempting to think that host immunity could be enlisted to enhance the clearance of drug resistant parasites, the biological mechanisms underlying this phenomenon remain unclear and require further study.

The Road Ahead: Concluding thoughts

Malaria researchers are uniquely poised to observe evolution in real-time, as parasites adapt and become resistant to ACTs. The current challenge is to try and understand ACT resistance before it becomes widespread, so that resistance mutations can be used as biomarkers to guide treatment policy, resistance containment, and malaria control efforts. Numerous studies aimed at identifying the genetic basis of ACT resistance have generated a long list of candidate genes; however, a causal link between these candidates and *in vivo* resistance has yet to be shown. Nonetheless, malaria researchers are well positioned to use genomics-era technologies in order to identify possible "road signs" along the way to ACT resistance.

Antimalarial drug use over the last half-century has maximized the evolutionary advantages of drug-resistant parasites. Synthetic compounds, given as monotherapy and aimed at eliminating every parasite in an infection, have driven the emergence and spread of resistant parasites. The recent shift in drug policy towards combination therapies is certainly a step in the right direction, but the lack of new antimalarial drugs on the horizon underscores the fact that once ACTs are rendered ineffective there will be few, if any, treatment options for patients that are sick with multi-drug resistant parasites. More diverse combination therapies are certainly needed; one recent study uncovered the antimalarial activities of numerous drugs already approved for treatment of other diseases, as well as compounds that appear to selectively target drug resistant parasites [26]. Collaborative efforts between researchers and drug companies have suggested the creation of a "malaria box" of publicly accessible compounds having antimalarial activity and would serve as a community resource [77]. Finally, drug resistance monitoring in naturally circulating parasite populations is essential, as the only way to know whether resistant parasites are in a particular area is to look for them [78]. These are just a few of the many efforts

underway to try and keep pace with parasites as they become increasingly drug resistant and cause infections that are more and more difficult for clinicians to manage.

Regardless of how it arises, it is clear that full-blown ACT resistance would be devastating for the millions of people stricken with malaria each year. In trying to identify and understand the genetic basis of resistance, it is important to keep an open mind. Drug resistance is a near-universal phenomenon among rapidly evolving human diseases, and researchers should seek applicable insights from other fields wherever possible. We hope that the years ahead will bring reliable markers of ACT resistance, a better understanding of how parasites become resistant, and new therapies that can combat resistant parasites. All of these will surely aid efforts to better manage drug-resistant malaria infections.

References

1. WHO (2011) World Malaria Report 2011

2. WHO (2010) *Guidelines for the treatment of malaria*. WHO Press

3. Dondorp, A.M.*, et al.* (2009) Artemisinin resistance in Plasmodium falciparum malaria. *NEJM* 361, 455-467

4. Paguio, M.F.*, et al.* (2011) Plasmodium falciparum resistance to cytocidal versus cytostatic effects of chloroquine. *Molecular and biochemical parasitology* 178, 1-6

5. Vestergaard, L.S. and Ringwald, P. (2007) Responding to the challenge of antimalarial drug resistance by routine monitoring to update national malaria treatment policies. *The American journal of tropical medicine and hygiene* 77, 153-159

6. Bloland, P.B. (2001) *Drug resistance in malaria*. WHO

7. Rieckmann, K.H.*, et al.* (1978) DRUG SENSITIVITY OF PLASMODIUM FALCIPARUM: An In-vitro Microtechnique. *Lancet* 311, 22-23

8. Desjardins, R.E.*, et al.* (1979) Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrobial agents and chemotherapy* 16, 710-718

9. Noedl, H.*, et al.* (2003) Malaria drug-sensitivity testing: new assays, new perspectives. *Trends in parasitology* 19, 175-181

10. Mackinnon, M.J. and Marsh, K. (2010) The selection landscape of malaria parasites. *Science* 328, 866-871

11. Fernandez, L.*, et al.* (2011) Creeping baselines and adaptive resistance to antibiotics. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* 14, 1-21

12. Mok, S.*, et al.* (2011) Artemisinin resistance in Plasmodium falciparum is associated with an altered temporal pattern of transcription. *BMC genomics* 12, 391

13. Garcia, J.E.*, et al.* (2006) Developmental biology of sporozoite-host interactions in Plasmodium falciparum malaria: implications for vaccine design. *Clinical microbiology reviews* 19, 686-707

14. Costanzo, M.S. and Hartl, D.L. (2011) The evolutionary landscape of antifolate resistance in Plasmodium falciparum. *Journal of genetics* 90, 187-190

15. Valderramos, S.G. and Fidock, D.A. (2006) Transporters involved in resistance to antimalarial drugs. *Trends in pharmacological sciences* 27, 594-601

16. Mita, T.*, et al.* (2009) Spread and evolution of Plasmodium falciparum drug resistance. *Parasitology international* 58, 201-209

17. Gardner, M.J.*, et al.* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419, 498-511

18. Ekland, E.H. and Fidock, D.A. (2007) Advances in understanding the genetic basis of antimalarial drug resistance. *Current opinion in microbiology* 10, 363-370

19. Volkman, S.K.*, et al.* (2012) Harnessing genomics and genome biology to understand malaria biology. *Nature reviews. Genetics* 13, 315-328

20. Ranford-Cartwright, L.C. and Mwangi, J.M. (2012) Analysis of malaria parasite phenotypes using experimental genetic crosses of Plasmodium falciparum. *International journal for parasitology* 42, 529-534

21. Wellems, T.E.*, et al.* (1991) Genetic mapping of the chloroquine-resistance locus on Plasmodium falciparum chromosome 7. *Proceedings of the National Academy of Sciences of the United States of America* 88, 3382-3386

22. Fidock, D.A.*, et al.* (2000) Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell* 6, 861-871

23. Beez, D.*, et al.* (2011) Genetic predisposition favors the acquisition of stable artemisinin resistance in malaria parasites. *Antimicrobial agents and chemotherapy* 55, 50-55

24. Mu, J.*, et al.* (2010) Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nature genetics* 42, 268-271

25. Van Tyne, D.*, et al.* (2011) Identification and Functional Validation of the Novel Antimalarial Resistance Locus PF10_0355 in Plasmodium falciparum. *PLoS Gen* 7, e1001383

26. Yuan, J.*, et al.* (2011) Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* 333, 724-729

27. Park, D.J.*, et al.* (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*

28. Manske, M.*, et al.* (2012) Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* 487, 375-379

29. Grossman, S.R.*, et al.* (2010) A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. *Science*

30. Voight, B.F., *et al.* (2006) A map of recent positive selection in the human genome. *PLoS Bio* 4, e72

31. Sabeti, P.C., *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913-918

32. Cheeseman, I.H., *et al.* (2012) A major genome region underlying artemisinin resistance in malaria. *Science* 336, 79-82

33. Nzila, A. and Mwai, L. (2010) In vitro selection of Plasmodium falciparum drug-resistant parasite lines. *The Journal of antimicrobial chemotherapy* 65, 390-398

34. Hunt, P., *et al.* (2010) Experimental evolution, genetic analysis and genome re-sequencing reveal the mutation conferring artemisinin resistance in an isogenic lineage of malaria parasites. *BMC genomics* 11, 499

35. Nam, T.G., *et al.* (2011) A chemical genomic analysis of decoquinate, a Plasmodium falciparum cytochrome b inhibitor. *ACS chemical biology* 6, 1214-1222

36. Mackinnon, M.J., *et al.* (2009) Comparative transcriptional and genomic analysis of Plasmodium falciparum field isolates. *PLoS pathogens* 5, e1000644

37. Milner, D.A., Jr., *et al.* (2012) Transcriptional profiling of Plasmodium falciparum parasites from patients with severe malaria identifies distinct low vs. high parasitemic clusters. *PloS one* 7, e40739

38. O'Brien, C., *et al.* (2011) Recent clinical and molecular insights into emerging artemisinin resistance in Plasmodium falciparum. *Current opinion in infectious diseases* 24, 570-577

39. Tucker, M.S., *et al.* (2012) Phenotypic and genotypic analysis of in vitro-selected artemisinin-resistant progeny of Plasmodium falciparum. *Antimicrobial agents and chemotherapy* 56, 302-314

40. Cui, L., *et al.* (2012) Mechanisms of in vitro resistance to dihydroartemisinin in Plasmodium falciparum. *Molecular microbiology* 86, 111-128

41. Imwong, M., *et al.* (2010) Exploring the contribution of candidate genes to artemisinin resistance in Plasmodium falciparum. *Antimicrobial agents and chemotherapy* 54, 2886-2892

42. White, N.J. and Pongtavornpinyo, W. (2003) The de novo selection of drug-resistant malaria parasites. *Proceedings. Biological sciences / The Royal Society* 270, 545-554

43. Wongsrichanalai, C., *et al.* (2002) Epidemiology of drug-resistant malaria. *The Lancet infectious diseases* 2, 209-218

44. Chait, R.*, et al.* (2012) What counters antibiotic resistance in nature? *Nature chemical biology* 8, 2-5

45. Hsu, E. (2006) The history of qing hao in the Chinese materia medica. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 100, 505-508

46. Ribacke, U.*, et al.* (2007) Genome wide gene amplifications and deletions in Plasmodium falciparum. *Molecular and biochemical parasitology* 155, 33-44

47. Nair, S.*, et al.* (2008) Adaptive copy number evolution in malaria parasites. *PLoS Gen* 4, e1000243

48. Cowman, A.F.*, et al.* (1994) Selection for mefloquine resistance in Plasmodium falciparum is linked to amplification of the pfmdr1 gene and cross-resistance to halofantrine and quinine. *Proceedings of the National Academy of Sciences of the United States of America* 91, 1143-1147

49. Sidhu, A.B.*, et al.* (2006) Decreasing pfmdr1 copy number in plasmodium falciparum malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *The Journal of infectious diseases* 194, 528-535

50. Foote, S.J.*, et al.* (1990) Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in Plasmodium falciparum. *Nature* 345, 255-258

51. Picot, S.*, et al.* (2009) A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria. *Malaria journal* 8, 89

52. Elde, N.C.*, et al.* (2012) Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses. *Cell* 150, 831-841

53. Pranting, M. and Andersson, D.I. (2011) Escape from growth restriction in small colony variants of Salmonella typhimurium by gene amplification and mutation. *Molecular microbiology* 79, 305-315

54. Andersson, D.I. (2003) Persistence of antibiotic resistant bacteria. *Current opinion in microbiology* 6, 452-456

55. Martinez-Picado, J. and Martinez, M.A. (2008) HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. *Virus research* 134, 104-123

56. Godin-Heymann, N.*, et al.* (2007) Oncogenic activity of epidermal growth factor receptor kinase mutant alleles is enhanced by the T790M drug resistance mutation. *Cancer research* 67, 7319-7326

57. Ecker, A.*, et al.* (2011) Evidence that mutant PfCRT facilitates the transmission to mosquitoes of chloroquine-treated Plasmodium gametocytes. *The Journal of infectious diseases* 203, 228-236

58. Bell, A.S.*, et al.* (2012) Enhanced transmission of drug-resistant parasites to mosquitoes following drug treatment in rodent malaria. *PloS one* 7, e37172

59. Sutherland, C.J.*, et al.* (2002) Gambian children successfully treated with chloroquine can harbor and transmit Plasmodium falciparum gametocytes carrying resistance genes. *The American journal of tropical medicine and hygiene* 67, 578-585

60. Osman, M.E.*, et al.* (2007) Field-based evidence for linkage of mutations associated with chloroquine (pfcrt/pfmdr1) and sulfadoxine-pyrimethamine (pfdhfr/pfdhps) resistance and for the fitness cost of multiple mutations in P. falciparum. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 7, 52-59

61. Mendez, F.*, et al.* (2007) Selection of antifolate-resistant Plasmodium falciparum by sulfadoxine-pyrimethamine treatment and infectivity to Anopheles mosquitoes. *The American journal of tropical medicine and hygiene* 77, 438-443

62. Mita, T.*, et al.* (2003) Recovery of chloroquine sensitivity and low prevalence of the Plasmodium falciparum chloroquine resistance transporter gene mutation K76T following the discontinuance of chloroquine use in Malawi. *The American journal of tropical medicine and hygiene* 68, 413-415

63. Hastings, I.M. and Donnelly, M.J. (2005) The impact of antimalarial drug resistance mutations on parasite fitness, and its implications for the evolution of resistance. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* 8, 43-50

64. Hayward, R.*, et al.* (2005) pfmdr1 mutations associated with chloroquine resistance incur a fitness cost in Plasmodium falciparum. *Molecular microbiology* 55, 1285-1295

65. Babiker, H.A. (2009) Seasonal fluctuation of drug-resistant malaria parasites: a sign of fitness cost. *Trends in parasitology* 25, 351-352

66. Hastings, I.M. and Watkins, W.M. (2006) Tolerance is the key to understanding antimalarial drug resistance. *Trends in parasitology* 22, 71-77

67. Witkowski, B.*, et al.* (2010) Increased tolerance to artemisinin in Plasmodium falciparum is mediated by a quiescence mechanism. *Antimicrobial agents and chemotherapy* 54, 1872-1877

68. Walliker, D.*, et al.* (2005) Fitness of drug-resistant malaria parasites. *Acta tropica* 94, 251-259

69. Brown, K.M.*, et al.* (2010) Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Molecular biology and evolution* 27, 2682-90

70. Kidgell, C.*, et al.* (2006) A systematic map of genetic variation in Plasmodium falciparum. *PLoS Path* 2, e57

71. Gonzales, J.M.*, et al.* (2008) Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Bio* 6, e238

72. Jiang, H.*, et al.* (2008) Genome-wide compensatory changes accompany drug- selected mutations in the Plasmodium falciparum crt gene. *PloS one* 3, e2484

73. Djimde, A.A.*, et al.* (2003) Clearance of drug-resistant parasites as a model for protective immunity in Plasmodium falciparum malaria. *The American journal of tropical medicine and hygiene* 69, 558-563

74. Enevold, A.*, et al.* (2007) Potential impact of host immunity on malaria treatment outcome in Tanzanian children infected with Plasmodium falciparum. *Malaria journal* 6, 153

75. Diakite, M.*, et al.* (2011) Host candidate gene polymorphisms and clearance of drug-resistant Plasmodium falciparum parasites. *Malaria journal* 10, 250

76. Nair, S.*, et al.* (2003) A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Molecular biology and evolution* 20, 1526-1536

77. Guiguemde, W.A.*, et al.* (2012) Global phenotypic screening for antimalarials. *Chemistry & biology* 19, 116-129

78. Sibley, C.H.*, et al.* (2008) A network to monitor antimalarial drug resistance: a plan for moving forward. *Trends in parasitology* 24, 43-48

CHAPTER ONE

Identification and Functional Validation of the Novel Antimalarial Resistance Locus *PF10_0355* in *Plasmodium falciparum*

Daria Van Tyne[1*], Daniel J. Park[2,3*], Stephen F. Schaffner[2*], Daniel E. Neafsey[2*], Elaine Angelino[4,5*], Joseph F. Cortese[2], Kayla G. Barnes[1], David M. Rosen[1], Amanda K. Lukens[1], Rachel F. Daniels[2,6], Danny A. Milner Jr.[1], Charles A. Johnson[2], Ilya Shlyakhter[2,3,4], Sharon R. Grossman[2,3,4,5], Justin S. Becker[1], Daniel Yamins[4], Elinor K. Karlsson[2,3,4], Daouda Ndiaye[7], Ousmane Sarr[7], Souleymane Mboup[7], Christian Happi[8], Nicholas A. Furlotte[9], Eleazar Eskin[9], Hyun Min Kang[10,] Daniel L. Hartl[3], Bruce W. Birren[2], Roger C. Wiegand[2], Eric S. Lander[2], Dyann F. Wirth[1,2*], Sarah K. Volkman[1,2,11*], Pardis C. Sabeti[2,3,4*]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

[2] Broad Institute, Cambridge, MA, USA

[3] Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

[4] FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA

[5] Harvard Medical School, Boston, MA, USA

[6] Biological and Biomedical Sciences, Harvard University, Cambridge, MA, USA

[7] Faculty of Medicine and Pharmacy, Cheikh Anta Diop University, Dakar, Senegal

[8] Malaria Research Laboratories, College of Medicine, University of Ibadan, Ibadan, Nigeria

[9] Department of Computer Science and Department of Human Genetics, University of California Los Angeles, Los Angeles, CA, USA

[10] Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

[11] School for Health Sciences, Simmons College, Boston, MA, USA

[*] These authors contributed equally to this work

CONTRIBUTIONS

This manuscript was published in *PLoS Genetics* in April 2011, and is reprinted here with permission. Supplemental information can be found in Appendix A. The data summarized in Figure 1.3 (*PF10_0355* overexpressing transgenic parasites), Figure 1.5A and 1.5C (*PF10_0355* copy number variation), Supplemental Table S6 (*PF10_0355* copy number quantification by qPCR) and Supplemental Figure S12 (*PF10_0355* copy number quantification by quantitative Southern blotting) were generated by Daria Van Tyne. The remaining figures were generated by the other co-first and final authors. The manuscript was written by Daria Van Tyne and the other co-first and final authors.

Abstract

The *Plasmodium falciparum* parasite's ability to adapt to environmental pressures, such as the human immune system and antimalarial drugs, makes malaria an enduring burden to public health. Understanding the genetic basis of these adaptations is critical to intervening successfully against malaria. To that end, we created a high-density genotyping array that assays over 17,000 single nucleotide polymorphisms (~1 SNP/kb), and applied it to 57 culture-adapted parasites from three continents. We characterized genome-wide genetic diversity within and between populations and identified numerous loci with signals of natural selection, suggesting their role in recent adaptation. In addition, we performed a genome-wide association study (GWAS), searching for loci correlated with resistance to thirteen antimalarials; we detected both known and novel resistance loci, including a new halofantrine resistance locus, *PF10_0355*. Through functional testing we demonstrated that *PF10_0355* overexpression decreases sensitivity to halofantrine, mefloquine and lumefantrine but not to structurally unrelated antimalarials, and that increased gene copy number mediates resistance. Our GWAS and follow-on functional validation demonstrate the potential of genome-wide studies to elucidate functionally important loci in the malaria parasite genome.

Author Summary

Malaria infection with the human pathogen, *Plasmodium falciparum*, results in almost a million deaths each year, mostly in African children. Efforts to eliminate malaria are underway, but the parasite is adept at eluding both the human immune response and antimalarial treatments. Thus, it is important to understand how the parasite becomes resistant to drugs and to develop strategies to overcome resistance mechanisms. Toward this end, we used population genetic

strategies to identify genetic loci that contribute to parasite adaptation and to identify candidate genes involved in drug resistance. We examined over 17,000 genetic variants across the parasite genome in over 50 strains in which we also measured responses to many known antimalarial compounds. We found a number of genetic loci showing signs of recent natural selection and a number of loci potentially involved in modulating the parasite's response to drugs. We further demonstrated that one of the novel candidate genes (*PF10_0355*) modulates resistance to the antimalarial compounds halofantrine, mefloquine and lumefantrine. Overall, this study confirms that we can use genome-wide approaches to identify clinically relevant genes and demonstrates through functional testing that at least one of these candidate genes is indeed involved in antimalarial drug resistance.

Introduction

Plasmodium falciparum malaria is a major public health challenge that contributes significantly to global morbidity and mortality. Efforts to control and eliminate malaria combine antimalarial drugs, bed nets and indoor residual spraying, with vaccine development a longer-term goal. Genetic variation in the parasite population threatens to undermine these efforts, as the parasite evolves rapidly to evade host immune systems, drugs and vaccines. Studying genetic variation in parasite populations will expand our understanding of basic parasite biology and its ability to adapt, and will allow us to track parasites as they respond to intervention efforts. Such understanding is increasingly important as countries move towards reducing disease burden and the ultimate elimination of malaria.

Given the potential impact of rapid evolution of *P. falciparum* in response to control and eradication strategies, discovery and characterization of *P. falciparum* genetic diversity has accelerated in recent years. Since the first malaria genome was sequenced in 2002 [1], over 60,000 unique SNPs have been identified by concerted sequencing efforts [2-4], and several genomic tiling arrays [5-9] and low-density SNP arrays [10, 11] have been developed to query this genetic variation. Recently the first malaria GWAS was published [11], in which 189 drug-phenotyped parasites from Asia, Africa and the Americas were genotyped using a low-density array (3,257 SNPs); that study identified loci under positive selection and found several novel drug resistance candidates.

For our study, we adopted two strategies for identifying genes involved in the malaria parasite's adaptive response: searching for signals of recent or ongoing natural selection, and searching for loci associated with one important clinical adaptation—resistance to antimalarial drugs. To make these searches possible, we began by sequencing 9 geographically diverse strains of *P. falciparum* to identify novel variation, thereby doubling the number of publicly available SNPs to 111,536 (all accessible at plasmodb.org), and used these SNPs to develop a high-density genotyping array assaying 17,582 validated markers. After characterizing linkage disequilibrium and population structure in our samples, we used the arrays to search for evidence of both ongoing balancing selection and recent positive selection, and to carry out a GWAS that sought loci associated with resistance to thirteen antimalarial agents. We then followed up one of the novel loci associated with drug resistance in order to verify that variation there was biologically involved in modulating drug response.

Materials and Methods

Parasites, Drug Testing, and DNA Isolation

Parasite samples and origins are detailed in Supplemental Methods and Table S1. Parasites were maintained by standard methods [12] and were tested for their response to amodiaquine, artemether, artesunate, artemisinin, atovaquone, chloroquine, dihydroartemisinin, halofuginone, halofantrine, lumefantrine, mefloquine, piperaquine and quinine according to the methods outlined by Baniecki, et al. [13] (Table S4, Fig. S13, Supplemental Methods). Follow-up drug testing was done by measuring uptake of ^{3}H-hypoxanthine [14]. Nucleic acids were obtained from parasite cultures using Qiagen genomic-tips (Qiagen, USA). All DNA samples were evaluated by molecular barcode [15].

Array Genotyping

We sequenced nine geographically diverse parasite isolates to 1.25x coverage, nearly doubling the number of publicly available SNPs to 111,536 (Supplementary Methods). These parasites had been previously sequenced to 0.25x coverage [2] and the deeper sequencing allowed for more thorough SNP discovery. Using this combined marker set, we created a high-density Affymetrix-based SNP array for *P. falciparum* containing 74,656 markers. Arrays were hybridized to 57 independent parasite samples (Table S1), including 17 previously sequenced strains used as a validation set. Genotype calls were produced using the BRLMM-P algorithm [16]. Markers that did not demonstrate perfect concordance between sequence and array data for the 17 strains were removed (Supplemental Methods). The remaining 17,582 SNPs constituted the high-confidence marker set used throughout this study (median marker spacing 444 bp, mean spacing 1,316 bp). All genomic positions and translation consequences are listed with respect to the PlasmoDB 5.0 assembly and annotation. SNP genotype data are publicly available on

plasmodb.org (release 6.0, July 2009) and dbSNP (Build B134, May 2011), accessible by searching for submission batches Pf_0002 (sequencing of nine isolates) and Pf_0003 (genotyping of 57 isolates) from submitter BROAD-GENOMEBIO.

Principal Component Analyses

Principal components analysis (PCA) was performed using the program SmartPCA [17]. All single-infection samples were used for the analysis in Figure 1.1. Samples that tightly clustered with the wrong continental population (A4, Malayan Camp and T2_C6) represented likely cases of contamination and were thus omitted from all other analyses.

Diversity/Divergence Analysis

We measured diversity using a statistic we term 'SNP π,' which quantifies the average number of pair-wise differences among samples from a given population at assayed SNPs. Population divergence was measured using $F_{ST,}$ calculated using the method of Hudson, et al. [18]. Statistical evaluation of the significance of differences in SNP π and F_{ST} among populations was performed using a bootstrapping approach, where the SNP set was re-sampled with replacement and each statistic recomputed 1000 times.

Linkage Disequilibrium (LD) Analysis

The statistic r^2 was calculated within each population for all pairs of SNPs sharing the same chromosome [19]; pairs were binned by distance and averaged within each bin. The level of LD between unlinked markers was estimated by calculating r^2 between all pairs of SNPs on different chromosomes. To determine the bias caused by small sample size, the unlinked calculation was repeated, with the change that for each pair of SNPs, the genotype for one was taken from one strain while the genotype for the second was taken from another strain. This

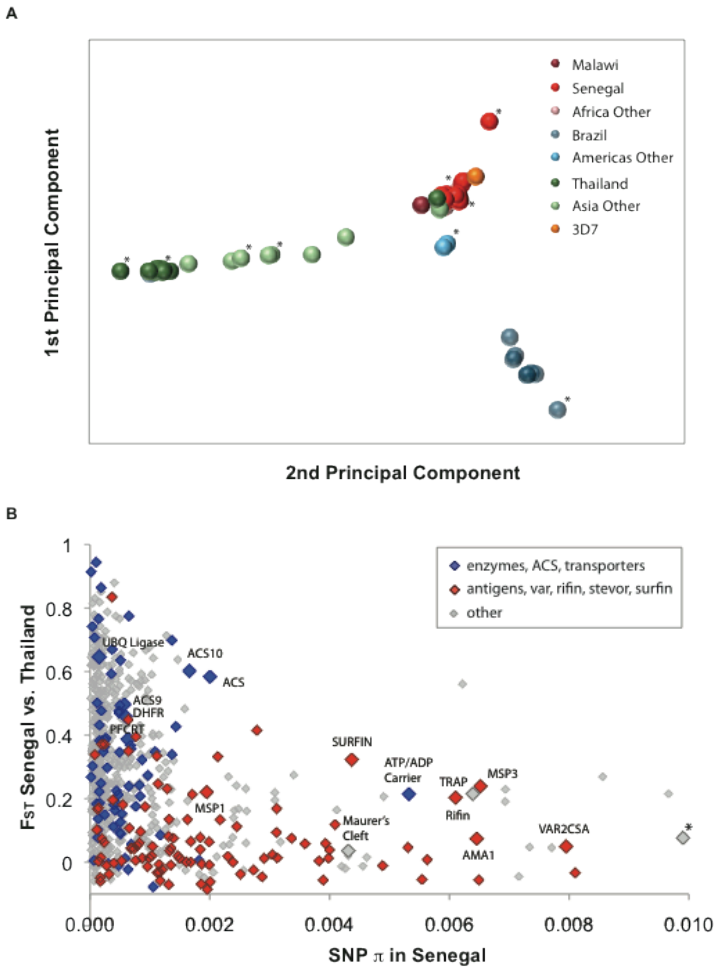


Figure 1.1. Parasite global population structure and genetic diversity vs. divergence. A) Population structure is visualized using the first two principal components of genetic variation for 57 parasites. Solid circles represent individual parasites, with colors assigned by reported origin: Africa in red, America in blue, and Asia in green. The nine strains used for ascertainment sequencing are indicated with (*). B) Genetic diversity (SNP p) in Senegal versus divergence (F_{ST}) between Senegal and Thailand is reported for 688 genes containing >3 successfully genotyped SNPs. Blue diamonds: enzymes, acyl-CoA synthetases (ACS) or transporters; red diamonds: antigens, vars, rifins, stevors or surfins; gray diamonds: all other genes. Gene IDs (PlasmoDB.org) for highlighted genes are listed in Table S7. A gene with unknown function is flagged with (*) to indicate that SNP π is off-scale (0.014).

background value of r^2 was calculated separately for the possible pairs of different strains and then averaged. Only single infections, as assessed by molecular barcode, were used.

Long Range Haplotype (LRH) Analysis

Because of the small number of samples, LRH results for individual continental populations had a high level of variance. Thus, we pooled together samples from Africa (n = 26) and Asia (n = 18, excluding India), as suggested by our PCA analysis. SNPs included in the analysis had a minor allele frequency of at least 0.05 and a call rate of at least 0.8; missing genotypes were imputed using PHASE. LRH analysis was performed using Sweep. Each SNP defined two core alleles, one base pair in length. We calculated relative extended haplotype homozygosity (REHH) for each core allele, to its left and right [20], yielding up to four REHH scores per SNP locus. We standardized the REHH scores as a function of core allele frequency, defined on a discrete grid from 0.05 to 0.95 with even spaces of 0.025. This yielded a normally-distributed set of *Z*-scores for which we calculated corresponding *P*-values and *Q*-values.

Genome Wide Association Study (GWAS)

We performed a GWAS for drug resistance to thirteen antimalarials across 50 of our genotyped samples. 7,437 SNPs that had a minor allele count of five samples as well as an 80% call rate under every phenotype condition were used for GWAS. A Bonferroni significance threshold of $-\log_{10}(P\text{-value}) > 5.17$ was used for all tests. See Supplemental Methods for more details on GWAS methods.

The Efficient Mixed-Model Association (EMMA) test [21] models quantitative trait associations to a data set with complex population structure and hidden relatedness. It calculates a genotype similarity matrix instead of discrete categories and does not require *a priori* specification of populations. The resulting *P*-value distributions demonstrate little remaining

effect from population structure (Fig. S8) while retaining power to find a number of associations at genome-wide significance (Fig. S8, 1.2A, Table 1.1).

The Haplotype Likelihood Ratio (HLR) test [22] models the likelihood that a single, resistant haplotype rose to dominance while all other haplotypes proportionally decreased. PLINK [23] is used to produce sliding window haplotypes across the genome and calculate haplotype frequencies for input to the HLR test. We produced input for all 2-, 4- and 6-marker windows. The LOD scores generated by the HLR test were converted to empirical pointwise *P*-values by performing approximately 370,000 permutations of the null model for each test condition, allowing us to calculate empirical *P*-values up to a significance of $10^{-5.6}$. We preserved population-specific phenotype frequencies by permuting only within each of three populations defined by our PCA analysis (Table S1). Resulting *P*-value distributions fit expectations well for the vast majority of test conditions (Fig. S9, S10) and the test demonstrates power to detect a number of loci at genome-wide significance (Fig. 1.2A, Table 1.1).

Copy Number Variation (CNV)

Copy number was assessed by evaluating the hybridization intensity at the *PF10_0355* locus on the high-density SNP array (Supplemental Methods). Follow-up analyses were done by quantitative real-time PCR (qPCR) of the *PF10_0355* locus using the Delta Delta Ct method [24]. *PF10_0355* was compared to the reference locus *PF07_0076* and 3D7 was used as a reference strain. A summary of *PF10_0355* copy number for all parasite strains tested is provided in Table S6. Select resistant strains that were found to have multiple copies of *PF10_0355* were further analyzed by quantitative Southern blotting and *PF10_0355* copy number was compared to the *dhps* gene from the 3D7 strain [25].

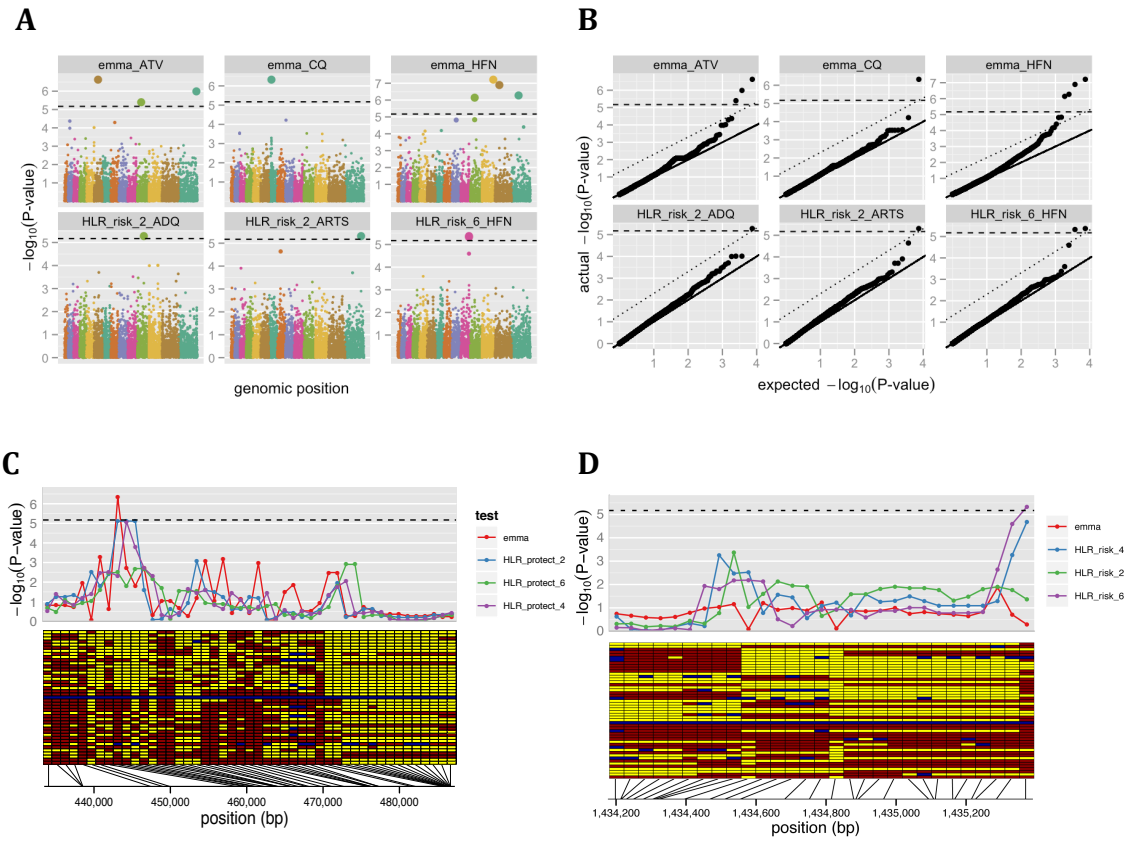


Figure 1.2. Genome-wide association study (GWAS) results. A) Genome-wide significant associations were found for five antimalarials (out of thirteen tested) using EMMA and HLR tests. They include *pfcrt* (chromosome 7) associated with chloroquine resistance and eleven novel associations with resistance to several drugs, listed in Table 1.1. B) Quantile-quantile plots for the *P*-value distributions in A) show no significant confounding from population structure. Bonferroni-corrected genome-wide significance is marked with a dashed line; Benjamini-Hochberg significance is marked with a dotted line. C-D) Close-ups are shown of the GWAS signal (top) and haplotypes (bottom) for resistance to C) chloroquine (CQ) around the gene *pfcrt* and D) halofantrine (HFN) around the gene *PF10_0355*. Yellow: sensitive allele; red: resistant allele; Blue: no data. Isolates are ordered by IC_{50}, with the highest IC_{50} on the bottom.

chr	SNPs	test	drug	*P*-value	genes	PlasmoDB description
6	674,154	EMMA	ATV	2.36E-07	PFF0785w	Ndc80 homologue, putative
7	459,787	EMMA	CQ	4.72E-07	MAL7P1_27	chloroquine resistance transporter
10	1,435,226, 1,435,286, 1,435,370, 1,437,695, 1,437,718, 1,441,590, 1,444,868	HLR_risk_6 (2 overlapping hits)	HFN	4.71E-06, 4.25E-06	PF10_0355, PF10_0356	erythrocyte membrane protein putative, liver stage antigen 1
11	657,349	EMMA	ATV	4.01E-06	PF11_0178	conserved unknown
11	738,407	EMMA	HFN	7.20E-07	PF11_0203	peptidase, putative
11	1,123,028, 1,124,030	HLR_risk_2	ADQ	5.26E-06	PF11_0302	conserved unknown
12	1,964,935	EMMA	HFN	6.15E-08	PFL2285c	conserved unknown
13	757,689	EMMA	HFN	1.28E-07	PF13_0101	conserved unknown
14	1,233,470	EMMA	HFN	5.32E-07	PF14_0293	conserved unknown
14	2,814,793, 2,815,714	HLR_risk_2	ARTS	4.90E-06	PF14_0654	aminophospholipid transporter, putative
14	3,130,449	EMMA	ATV	1.03E-06	PF14_0729	early transcribed membrane protein 14.2

Table 1.1. Eleven genome-wide significant associations with antimalarial drug resistance. Positions are given with respect to the PlasmoDB 5.0 reference assembly of 3D7. Drug abbreviations are ATV: atovaquone; CQ: chloroquine; HFN: halofantrine; ADQ: amodiaquine; ARTS: artemisinin. The HLR test for CQ-*pfcrt* association is just below the genome-wide significance threshold and is omitted here, but is shown in Figure 1.2C.

PF10_0355 Overexpression

The full length ORF of *PF10_0355* was amplified from either the Dd2 (HFN sensitive) or SenP08.04 (HFN resistant) parasite isolate and cloned into the pBIC009 plasmid under the expression of the *Hsp86* promoter. Plasmid DNA was isolated, tranfected into the Dd2 parasite strain and stable transfectants were selected with 2.5nM WR99210 [26]. Parasites from two independent experiments for each vector type (Dd2+Dd2 and Dd2+SenP08.04) were isolated and successful transfection was confirmed by plasmid rescue as well as episome-specific PCR and sequencing. Additionally, a vector control strain was made by transfecting Dd2 parasites with the pBIC009 plasmid containing the firefly luciferase gene (EC 1.13.12.7).

Results

Genetic Diversity

We identified global population structure among malaria parasites using principal components analysis (PCA) of 57 genotyped culture-adapted parasite samples (Fig. 1.1A, Table S1, Fig. S1). African, American and Asian samples form three distinct clusters, reflecting the likely independent introduction of *P. falciparum* from Africa into Asia and the Americas. There was little evidence for structure within Africa, suggesting high gene flow throughout the region (Fig. S1). Asian and American parasites however both show substantial internal structure.

There are also dramatic differences in linkage disequilibrium (LD) between populations, with substantial LD extending less than 1 kb in Senegal, 10 kb in Thailand, and 100 kb in Brazil (Fig. S2). These observations are consistent with previous findings, which showed that LD decays more rapidly in Africa, due either to founder effects in other continents [27] or to elevated outcrossing frequencies in Africa [27, 28], where higher transmission intensity leads to

a greater likelihood of sexual outcrossing rather than selfing within the mid-gut of vector mosquitoes.

The short LD in malaria, driven by high levels of recombination, means that a high density of markers is required to identify candidate loci in association studies, since causal variants not on the array can seldom be tagged by neighboring alleles (Table S2). On the other hand, short LD can aid in fine-mapping candidate associations and greatly accelerates the search for causal genes. Short LD also aids in identifying genomic regions under recent positive selection with recombination-based methods, since the increased LD in selected regions should stand out against the short-LD background.

Detecting Signals of Natural Selection

We expect that many parasite proteins that interact with the host immune system will be under balancing selection, because they will be under selective pressure to maintain high levels of diversity. Indeed, previous studies have shown that regions of the *P. falciparum* genome that are highly polymorphic and appear to be under balancing selection encode antigens that are recognized by the human immune system [4]. We examined evidence for balancing selection in our data by searching for regions with high nucleotide diversity (as measured by SNP π) and low population divergence (as measured by F_{ST}) (Fig. 1.1B). When we examined the loci lying in this region of the graph (Fig. S3), we found a number of known antigens and vaccine candidates. Loci in the same region with unknown function are thus potential novel antigens that trigger human immune response to malaria, and may prove useful as biomarkers or as candidate vaccine molecules.

We carried out a similar search for loci under positive selection by identifying regions with both low nucleotide diversity within Senegal and Thailand and high population divergence between the two populations (Fig. 1.1B). We observed throughout the genome that nucleotide diversity was lower for nonsynonymous SNPs than for intergenic SNPs (Fig. S4), a characteristic result of widespread purifying selection. At the same time, nonsynonymous SNPs exhibited significantly greater divergence than intergenic SNPs in all pairwise population comparisons, suggesting the effect of positive selection in local *P. falciparum* populations. Nonsynonymous SNPs with low diversity within a population and high divergence between the two populations studied may represent polymorphisms responsible for adaptive evolution.

We also carried out a genome-wide scan for recent positive selection using the long-range haplotype (LRH) test [29], which identifies common variants that have recently spread to high prevalence using recombination as a clock. Approximately 15 genes were identified as having undergone recent positive selection by this approach, including known drug resistance loci (*pfcrt* and *dhfr*) as well as multiple members of the acyl-CoA synthetase (ACS) and ubiquitin protein ligase families (Fig. S5 and S6); these latter genes also exhibit high divergence between Senegal and Thailand (Fig. 1.1B), evidence for selection that is recent and population-specific. Taken as a group, the genes identified by the LRH test show a significant enrichment for higher than average population divergence (as measured by F_{ST}, Mann-Whitney U=1583, P=0.0071). All of these loci (Table S3), which include genes in the folate metabolism, lipid biosynthesis and ubiquitin pathways, should be viewed as candidates for functional follow-up and further characterization.

Genome-wide Associations with Drug Resistance

In order to directly assess the genetic basis for one important response to antimalarial intervention, we carried out a GWAS to identify loci associated with drug resistance in parasites. This same approach can potentially be applied to many other clinically relevant malaria phenotypes, e.g. host response, invasion, and gametocyte formation. Our first step was to measure drug resistance (IC_{50} values) to 13 antimalarial drugs (amodiaquine, artemether, artesunate, artemisinin, atovaquone, chloroquine, dihydroartemisinin, halofuginone, halofantrine, lumefantrine, mefloquine, piperaquine and quinine) in 50 culture-adapted parasites using a high-throughput assay (Table S4-5, Supplementary Methods).

We performed the genome-wide association analysis using two statistical tests: efficient mixed-model association (EMMA) and a haplotype likelihood ratio (HLR) test (Figs. S7-10, Methods). EMMA identifies quantitative trait associations in individuals with complex population structure and hidden relatedness; it has previously been shown to outperform both PCA-based and λ_{GC}-based correction approaches in highly inbred and structured mouse, maize, and Arabidopsis populations [21]. EMMA is particularly applicable for small and structured sample sets: one of its first applications was in a study of 24 diploid mouse strains [21], essentially the same sample size as in our study (50 haploid strains). The HLR test is a multi-marker test designed to detect the association of a single haplotype with a phenotype, and is particularly powerful when the associated haplotype experienced recent strong selection (and is therefore long) and occurs on a low-LD background [22]; it is therefore particularly appropriate for this study. We addressed the effect of population structure in the HLR test using population-specific permutation (Methods). When used together, these two complementary approaches provide a highly sensitive screen for association signals within the *P. falciparum* genome.

The well-characterized chloroquine resistance locus, *pfcrt,* served as a positive control for our GWAS methods (Fig. 1.2A,C, Table S2), an important test given our small sample size and the limited LD present in *P. falciparum*. As expected, we found evidence for association with resistance to chloroquine using both tests, consistent with previous studies [11]; EMMA yielded evidence for association with genome-wide signficance, while the signal from the HLR test fell just short of genome-wide significance (Fig. 1.2C).

Applying the same tests to the other drug phenotypes, we detected numerous novel loci showing significant associations with drug resistance (Fig. 1.2A,D, Table 1.1). Quantile-quantile plots for each test demonstrate that we were able to effectively control for population structure (Fig. 1.2B). Despite our small sample size and the low LD in *P. falciparum*, in total eleven loci achieved genome-wide significance for association with resistance to five different drugs: amodiaquine, artemisinin, atovaquone, chloroquine and halofantrine. In most cases, the short extent of LD allowed localization to individual genes. Among the loci identified were various transporters and membrane proteins, as well as five conserved genes with unknown function (Table 1.1).

Functional Validation of a Novel Resistance Candidate

Demonstrating that a signal of association actually reflects a causal molecular process requires functional testing and validation of the candidate locus, both because of concerns about power and reproducibility of genetic association tests, and because even a robust statistical correlation need not imply biological causation. To confirm the ability of GWAS to identify functionally relevant candidates, we investigated one of our association findings, *PF10_0355,* in greater depth. This gene contains multiple SNPs associated with halofantrine resistance (Fig.

1.2D), and encodes a putative erythrocyte membrane protein (PlasmoDB.org) characterized by high genetic diversity.

We set out to determine the role of *PF10_0355* in halofantrine resistance by transfecting halofantrine-sensitive Dd2 parasites with episomal plasmids containing the *PF10_0355* gene from a halofantrine-resistant parasite (SenP08.04), a technique that is used routinely for stable transgene expression [30]. Two independent transfectants overexpressing the *PF10_0355* gene from SenP08.04 both showed reduced susceptibility to halofantrine when compared with the Dd2 parent or a transfection control (Fig. 1.3A), suggesting that this gene is indeed involved in modulating parasite drug response.

Two independent transfectants overexpressing the endogenous *PF10_0355* gene from halofantrine-sensitive Dd2 also showed reduced susceptibility to halofantrine (Fig. 1.3A), however, pointing to a role of overexpression in the observed resistance. Because *PF10_0355* is annotated as a putative erythrocyte membrane protein and belongs to the merozoite surface protein 3/6 family, we tested the hypothesis that the observed effect was the by-product of a growth or invasion-related process, rather than resistance due to a direct interaction with the antimalarial itself. To that end, we expanded our drug testing in the transfectant lines to include other antimalarials, some structurally related and some unrelated to halofantrine.

Overexpression of *PF10_0355* from either the Dd2 or the SenP08.04 parent caused increased resistance to the structurally related antimalarials mefloquine and lumefantrine (Fig. 1.3B,C), but had no effect on parasite susceptibility to the structurally unrelated antimalarials chloroquine, artemisinin or atovaquone (Fig. 1.3D,E). Indeed, we found evidence of cross-resistance between halofantrine and both mefloquine and lumefantrine (Fig. 1.4). We also observed cross-resistance between halofantrine and artemisinin, which is expected as cross-

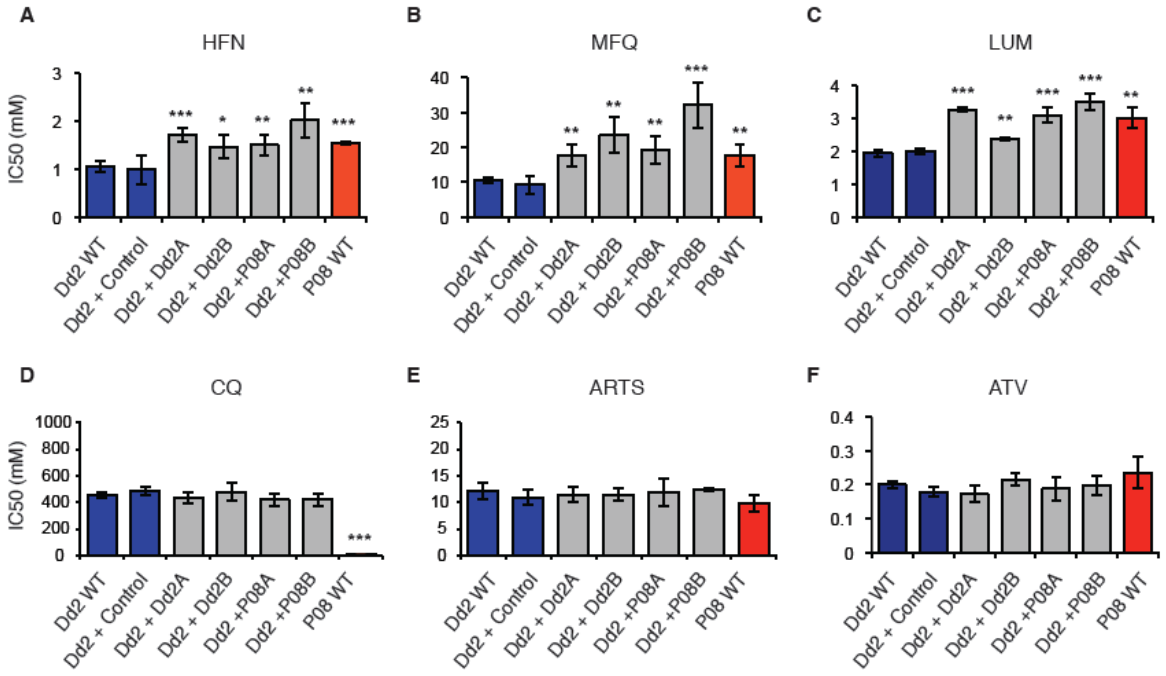


Figure 1.3. Overexpression of *PF10_0355* decreases parasite susceptibility to halofantrine (HFN) and related antimalarials. Parasite susceptibility to six antimalarials was measured by ^{3}H-hypoxanthine incorporation. Comparisons were made between Dd2 (HFN-sensitive strain) and SenP08.04 (HFN-resistant strain), as well as 4 transfected lines. "Dd2+Dd2": Dd2 parasites overexpressing *PF10_0355* from Dd2; "Dd2+P08": Dd2 parasites overexpressing *PF10_0355* from SenP08.04. Overexpression of *PF10_0355* decreases parasite susceptibility to A) HFN and structurally related B) mefloquine (MFQ) and C) lumefantrine (LUM). Overexpression of *PF10_0355* does not alter parasite susceptibility to D) chloroquine (CQ), E) artemisinin (ARTS) or F) atovaquone (ATV). Mean IC_{50} ± standard error is shown. Significance levels: *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

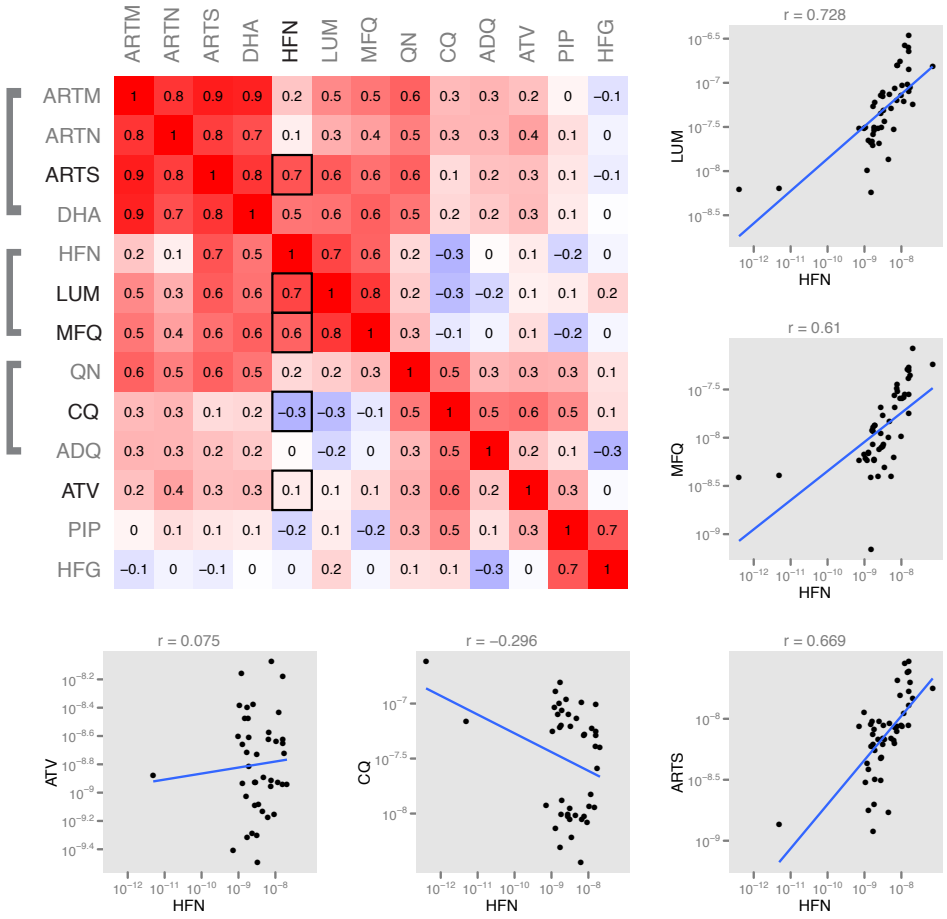


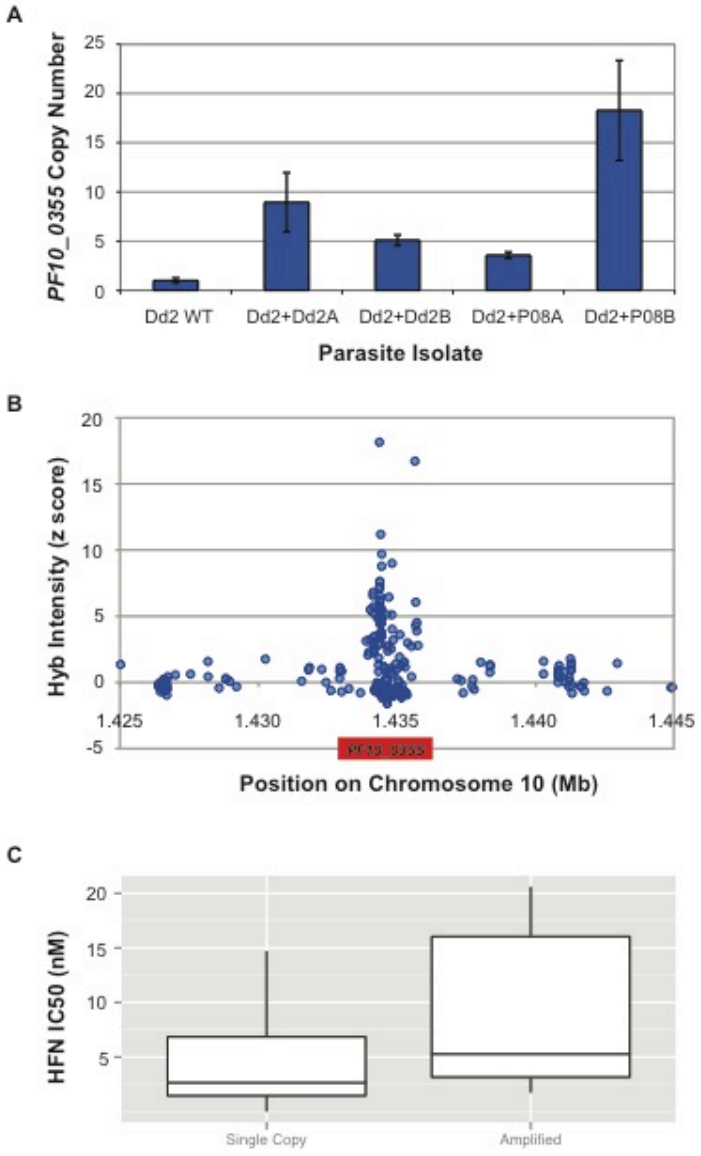
Figure 1.4. Correlations between antimalarial drugs tested. Pearson correlation values (r) between $\log_{10}(IC_{50})$ values are rendered as a color in a symmetric correlation matrix (red: correlated; white-uncorrelated, blue: inversely correlated). Thirteen antimalarials are measured: artemether (ARTM), artesunate (ARTN), artemisinin (ARTS), dihydroartemisinin (DHA), halofantrine (HFN), lumefantrine (LUM), mefloquine (MFQ), quinine (QN), chloroquine (CQ), amodiaquine (ADQ), atovaquone (ATV), piperaquine (PIP), and halofuginone (HFG). Drugs are grouped by structural relatedness. Correlation plots are given with a linear regression line for HFN compared to the 5 other drugs tested for antimalarial resistance with *PF10_0355* overexpression: (B) LUM, (C) MFQ, (D) ATV, (E) CQ, and (F) ARTS.

resistance between aminoquinolines and artemisinin compounds has been previously demonstrated [11, 31] and resistance to all these drugs has been shown to be mediated by changes in *pfmdr1* copy number [32, 33]. Overexpression of *PF10_0355*, however, alters parasite susceptibility to the aminoquinolines but not to artemisinin, suggesting that this effect is specific for that set of structurally related compounds and distinct from the effect of *pfmdr1*, which seems to exert a global effect of resistance to unrelated compounds (i.e. both aminoquinolines and artemisinins). Using the Dd2 parasite line, which has amplified *pfmdr1* copy number, as a background for *PF10_0355* overexpression allowed us to distinguish between cross-resistance to a structurally related class of compounds (mediated by *PF10_0355* overexpression) and pan-resistance to multiple classes of drugs.

Given that overexpression of the *PF10_0355* gene both from a halofantrine-resistant and from a sensitive parasite conferred resistance to halofantrine-related drugs, we investigated whether gene amplification might be driving the observed resistance, as it often does for antimalarial drugs [34-39]. We quantified *PF10_0355* copy number in our transfectants and found that the transfectant with the highest IC_{50} for all three drugs (Dd2+P08B) also had the highest *PF10_0355* copy number, as measured by quantitative PCR (qPCR) (Fig. 1.5A). Furthermore, when we examined the *PF10_0355* gene on our SNP array, we detected a substantial increase in hybridization intensity at the *PF10_0355* locus compared to the genome average, suggesting that this gene is amplified in some parasites (Fig. 1.5B). The amplified region appears only to contain the *PF10_0355* gene itself and not surrounding loci. We observed a similar pattern at *pfmdr1* on chromosome 5, where copy number variation is well established (Fig. S11). Follow-up qPCR analysis of 38 parasite lines confirmed that parasites with amplified

Figure 1.5. Copy number variation at *PF10_0355* is associated with HFN resistance. A) Mean *PF10_0355* copy number (± standard deviation for three replicates) in the parent Dd2 and transfected lines from qPCR analysis. Dd2+Dd2: Dd2 parasites overexpressing *PF10_0355* from HFN-sensitive Dd2; Dd2+P08: Dd2 parasites overexpressing *PF10_0355* from HFN-resistant SenP08.04. Copy number was compared to the reference locus *PF07_0076*. B) Increased hybridization intensity at *PF10_0355* on the high-density SNP array, measured by *Z*-scores for normalized and background-corrected data, for the HFN-resistant isolate SenP19.04. C) Strains with increased copy number of *PF10_0355* (as measured by qPCR > 1.2x 3D7) show a significantly higher resistance to HFN ($p = 0.02$, Student t-test).

Figure 1.5 (Continued)



PF10_0355 have a greater mean halofantrine IC_{50}. (Fig. 1.5C, Table S6). Copy number variation was further confirmed in a number of parasites by quantitative Southern blotting (Fig. S12).

Discussion

In this study we used natural selection and genome-wide association methods to probe the genetic basis of adaptation in *P. falciparum*. These approaches are complementary: scanning for selected loci permits an unbiased search for unknown adaptive changes, but provides little information about the processes at work, while GWAS gives a focused look at one easily identified (and clinically critical) adaptive phenotype. Results from both approaches open up new avenues for study, as we seek to understand the biological significance of the findings.

The specifics of our strategy were designed to cope with two potential limitations in applying genome-wide population genetic approaches to malaria: small sample sizes, due to the difficulty in adapting parasites to culture and assessing drug and other phenotypes; and a lack of correlation (LD) between nearby variants in the parasite genome, which limits our ability to infer untyped SNPs from genotyped markers. The second limitation we addressed by developing a high-density genotyping array (based on new sequencing), to increase the fraction of genetic variation that we could directly interrogate, while the effect of the first was mitigated by the phenotype we targeted in our GWAS.

Drug resistance is a phenotype well-suited for GWAS because it is expected to be caused by common alleles of large effect at few genomic loci [40]. If this is the case, associations will be much easier to detect than in a typical human GWAS, in which the phenotype is caused by alleles at many loci that are either rare or of small effect. Additionally, the haploid nature of the intra-erythrocytic stage of *P. falciparum* further heightens GWAS power by eliminating the issue

of allelic dominance. Finally, the increased LD caused by recent selection for drug resistance counteracts the loss of power that comes from short LD, small sample size, and the temporal and geographic stratification of the parasite population that we examined. Thus, despite the potential limitations, we were able to detect a known drug resistance locus (*pfcrt*), observed little *P*-value inflation in our GWAS data (Fig. S8, S9, S10), and identified a number of genome-wide significant loci associated with drug resistance. Part of this success was likely due to specific tests we used to account for population structure.

Going beyond these statistical tests, we went on to functionally validate one of these loci, demonstrating that increased *PF10_0355* copy number confer resistance to three structurally related antimalarial drugs. This demonstrates the feasibility of coupling GWAS and functional testing in the malaria parasite for identifying and validating novel drug resistance loci and illustrates the power of GWAS to find functionally important alleles.

Comparing our results to the recent GWAS described by Mu, et al. [11], which was also directed at finding drug-resistance loci, we see that, beyond the well-known *pfcrt* locus, there was no overlap between the associations identified by each study. Differing sets of drugs tested and analytical methods explain much of the disagreement. Of the eleven candidate associations in Table 1.1, one (that with *pfcrt*) was found by both studies, eight were associations with drugs not assayed in Mu, et al. (atovaquone and halofantrine), and two were found only with a haplotype-based test, an approach not used by Mu, et al. Our candidate locus at *PF10_0355*, in fact, would not have been detectable in the Mu study because it was identified only by the multi-marker HLR test, because it involved an association with halofantrine, and because the Mu, et al. genotyping array lacked markers within 4 kb of the gene (plasmoDB.org).

Different parasite populations and marker sets probably explain many of the dihydroartemisinin, mefloquine and quinine associations identified by Mu, et al. but not seen in our data set. The studies used different parasite population sets—theirs was weighted toward southeast Asian strains, and ours toward African strains—and selection pressures and selected alleles can both vary between populations. Our smaller sample size also means that we might lack power to identify some associations accessible to Mu, et al. These difficulties are reflected in human GWAS studies as well, where the ability to replicate associations using multiple tests and in different sample sets has also been challenging to achieve [41].

Ultimately, the disparities in loci identified point to the role of population analysis as a tool for candidate gene discovery and not as a definitive study. Even within each study, there is little overlap between the signals observed with different methods—our study detects only one gene (*pfcrt*) by both GWAS tests (EMMA and HLR), while Mu, et al. detected only two genes (unknowns, not *pfcrt*) by both of their GWAS tests (Eigensoft and PLINK). Even a well-designed GWAS serves only as a hypothesis-generating experiment, and it is vital to empirically validate candidate loci associated with a phenotype of interest. Especially given the small sample sizes and relatively sparse marker density used in both malaria GWAS studies to date, functional validation of candidates is necessary to address concerns about false positive results.

Our functional result, that increased *PF10_0355* copy number confers decreased susceptibility to halofantrine, mefloquine and lumefantrine, raises additional questions for study. Further work will be needed to determine the precise contributions of copy number variation and gene mutation to the parasite's response to these drugs. The biological function of this gene's product is unknown, but previous work indicates putative localization to the parasite surface [42], as well as it being a potential target of host immunity and balancing selection [43]. While

the protein itself does not appear to be a transporter, it is possible that it directly binds drug or perhaps couples with transport proteins to modulate drug susceptibility; interaction between membrane transporters and non-channel proteins has been demonstrated in cancer, plant and yeast systems [44-46]. Additional experiments are certainly required to determine the precise role of *PF10_0355* in modulating parasite response to this class of compounds, including assessing its relevance to resistance in natural populations, but it is clear that alteration of this locus can mediate drug resistance in *P. falciparum*.

Although halofantrine, mefloquine and lumefantrine are not commonly used as primary interventions, widespread halofantrine use has recently been documented in West Africa. Notably, halofantrine was used to treat nearly 18 million patients between 1988 and 2005 [47, 48], and it remains in production and use today. Use of halofantrine, mefloquine or lumefantrine as monotherapy may further explain how mutations and copy number variation in the *PF10_0355* gene were selected. Lumefantrine is also currently used as a partner drug in the artemisinin-based combination therapy (ACT) Coartem. The shorter half-life of artemether allows lumefantrine to be present as monotherapy, making it vulnerable to selection of drug resistant mutants. As genetic loci associated with drug responses are identified and validated, these provide new molecular biomarkers to evaluate drug use and response in malaria endemic settings. Thus, our findings have implications for defining molecular biomarkers for monitoring partner drug responses as intervention strategies, such as ACTs, are applied.

Beyond identifying a novel drug resistance locus, this study illustrates the general utility of a GWAS approach for the discovery of gene function in *P. falciparum*. Even with a small and geographically heterogeneous sample of parasites, we identified a number of new loci associated with drug response and validated one of them. Larger samples from a single population will have

much greater power to detect additional loci, including those where multiple and low frequency alleles contribute to resistance. Future GWAS have the potential both to provide greater insights into basic parasite biology and to identify biomarkers for drug resistance and other clinically relevant phenotypes like acquired protection, pathogenesis, and placental malaria.

Future GWAS will be able to counteract the loss of power caused by low LD, either by focusing on parasite populations with reduced outcrossing rates, or by studying cases of very strong selective pressure. This issue will soon become moot, however, as the declining cost of whole-genome sequencing makes it practical to assay every nucleotide in the genome on a routine basis. Culture-adapted parasites are amenable to robust and reproducible phenotypic characterization, but their limitations—the potential for artifactual mutations during adaptation and for a biased selection of clones within a given infection—mean that genetic changes identified using them require both functional validation and demonstration that the changes are important during natural infection. As direct sequencing of clinical isolates with demonstrable clinical phenotypes such as *ex vivo* drug response or invasion properties becomes increasingly feasible, sequencing will enable us to directly identify genetic changes in the parasite associated with clinically relevant phenotypes. In the years ahead, genome analysis of *P. falciparum* has the potential to identify genetic loci associated with many phenotypes, enhance our understanding of the biology of this important human pathogen, and inform the development of diagnostic and surveillance tools for malaria eradication.

Acknowledgments

The study is supported by the Bill and Melinda Gates Foundation, Ellison Medical Foundation, ExxonMobil Foundation, NIH Fogarty and NIAID. PCS is also supported by

fellowships from the Burroughs Wellcome and Packard Foundations. DJP is supported by an NSF Graduate Research Fellowship. We gratefully acknowledge B. Coleman, J. Dvorin, M.T. Duraisingh, U. Ribacke and C. Valim for help with overexpression vectors and useful discussions. T. Burke, N. Mahesh, G. Ramirez, and N. Senaratne provided technical help. Parasites lines or samples were provided by: J. Barnwell, A.P. Dash, C.E. Chitnis, K. Day, A. Djimde, C. Plowe, A.M. Katzin, D. Kyle, S. Thaithong, S.d.L. Moraes, J. Smith; and X. Su. Malaria Research and Reagent Resource Repository provided parasites deposited by: W. E. Collins, D.E. Kyle, L. H. Miller, D. Baruch, W. Trager, D. Walliker, U. Certa, R. Reber-Liske, T.E. Wellems, and Y. Wu (Supplementary Methods).

References

1. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. pp. 498-511.

2. Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, et al. (2007) A genome-wide map of diversity in Plasmodium falciparum. Nat Genet 39: 113-119.

3. Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, et al. (2007) Genome variation and evolution of the malaria parasite Plasmodium falciparum. Nat Genet. pp. 120-125.

4. Mu J, Awadalla P, Duan J, McGee KM, Keebler J, et al. (2007) Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. Nat Genet. pp. 126-130.

5. Carret CK, Horrocks P, Konfortov B, Winzeler E, Qureshi M, et al. (2005) Microarray-based comparative genomic analyses of the human malaria parasite Plasmodium falciparum using Affymetrix arrays. Mol Biochem Parasitol. pp. 177-186.

6. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, et al. (2006) A systematic map of genetic variation in Plasmodium falciparum. PLoS Pathog. pp. e57.

7. Jiang H, Yi M, Mu J, Zhang L, Ivens A, et al. (2008) Detection of genome-wide polymorphisms in the AT-rich Plasmodium falciparum genome using a high-density microarray. BMC Genomics. pp. 398.

8. Dharia NV, Sidhu AB, Cassera MB, Westenberger SJ, Bopp SE, et al. (2009) Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in Plasmodium falciparum. Genome Bio. pp. R21.

9. Tan JC, Patel JJ, Tan A, Blain JC, Albert TJ, et al. (2009) Optimizing comparative genomic hybridization probes for genotyping and SNP detection in Plasmodium falciparum. Genomics 93: 543-550.

10. Neafsey DE, Schaffner SF, Volkman SK, Park DJ, Montgomery P, et al. (2008) Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Bio. pp. R171.

11. Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, et al. (2010) Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nat Genet 42: 268-271.

12. Mu J, Awadalla P, Duan J, McGee KM, Joy DA, et al. (2005) Recombination hotspots and population structure in Plasmodium falciparum. PLoS Biol 3: e335.

13. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, et al. (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Mol Biol Evol. pp. 1467-1482.

14. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832-837.

15. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709-1723.

16. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803-819.

17. Crabb BS, Triglia T, Waterkeyn JG, Cowman AF (1997) Stable transgene expression in Plasmodium falciparum. Mol Biochem Parasitol 90: 131-144.

18. Pradines B, Hovette P, Fusai T, Atanda HL, Baret E, et al. (2006) Prevalence of in vitro resistance to eleven standard or new antimalarial drugs among Plasmodium falciparum isolates from Pointe-Noire, Republic of the Congo. J Clin Microbiol 44: 2404-2408.

19. Sidhu AB, Uhlemann AC, Valderramos SG, Valderramos JC, Krishna S, et al. (2006) Decreasing pfmdr1 copy number in plasmodium falciparum malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. J Infect Dis 194: 528-535.

20. Chavchich M, Gerena L, Peters J, Chen N, Cheng Q, et al. (2010) Role of pfmdr1 amplification and expression in induction of resistance to artemisinin derivatives in Plasmodium falciparum. Antimicrob Agents Chemother 54: 2455-2464.

21. Foote SJ, Thompson JK, Cowman AF, Kemp DJ (1989) Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of P. falciparum. Cell 57: 921-930.

22. Wilson CM, Serrano AE, Wasley A, Bogenschutz MP, Shankar AH, et al. (1989) Amplification of a gene related to mammalian mdr genes in drug-resistant Plasmodium falciparum. Science 244: 1184-1186.

23. Price RN, Uhlemann AC, Brockman A, McGready R, Ashley E, et al. (2004) Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. Lancet 364: 438-447.

24. Nair S, Miller B, Barends M, Jaidee A, Patel J, et al. (2008) Adaptive copy number evolution in malaria parasites. PLoS Genet 4: e1000243.

25. Anderson TJ, Patel J, Ferdig MT (2009) Gene copy number and malaria biology. Trends Parasitol 25: 336-343.

26. Ribacke U, Mok BW, Wirta V, Normark J, Lundeberg J, et al. (2007) Genome wide gene amplifications and deletions in Plasmodium falciparum. Mol Biochem Parasitol. pp. 33-44.

27. Hayton K, Su XZ (2008) Drug resistance and genetic mapping in Plasmodium falciparum. Curr Genet 54: 223-239.

28. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33: 177-182.

29. Singh S, Soe S, Weisman S, Barnwell JW, Perignon JL, et al. (2009) A conserved multi-gene family induces cross-reactive antibodies effective in defense against Plasmodium falciparum. PLoS One 4: e5410.

30. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, et al. (2010) Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in Plasmodium falciparum. Mol Biol Evol.

31. Miletti-Gonzalez KE, Chen S, Muthukumaran N, Saglimbeni GN, Wu X, et al. (2005) The CD44 receptor interacts with P-glycoprotein to promote cell migration and invasion in cancer. Cancer Res 65: 6660-6667.

32. Geisler M, Girin M, Brandt S, Vincenzetti V, Plaza S, et al. (2004) Arabidopsis immunophilin-like TWD1 functionally interacts with vacuolar ABC transporters. Mol Biol Cell 15: 3393-3405.

33. Beese SE, Negishi T, Levin DE (2009) Identification of positive regulators of the yeast fps1 glycerol channel. PLoS Genet 5: e1000738.

34. (2005) Halofantrine and fatal cardiac arrhythmia. GSK: Global Clinical Safety and Pharmacovigilance.

35. Bouchaud O, Imbert P, Touze JE, Dodoo AN, Danis M, et al. (2009) Fatal cardiotoxicity related to halofantrine: a review based on a worldwide safety data base. Malar J 8: 289.

36. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. Science 193: 673-675.

37. Baniecki ML, Wirth DF, Clardy J (2007) High-throughput Plasmodium falciparum growth assay for malaria drug discovery. Antimicrob Agents Chemother 51: 716-723.

38. Webster HK, Boudreau EF, Pavanand K, Yongvanitchit K, Pang LW (1985) Antimalarial drug susceptibility testing of Plasmodium falciparum in Thailand using a microdilution radioisotope method. Am J Trop Med Hyg 34: 228-235.

39. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, et al. (2008) A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. Malar J 7: 223.

40. (2007) BRLMM-P: a Genotype Calling Method for the SNP 5.0 Array http://www.affymetrix.com/support/technical/whitepapers/brlmmp_whitepaper.pdf.

41. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2: e190.

42. Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583-589.

43. Hill WG, Robertson A (1968) Linkage Disequilibrium in Finite Populations. Theoretical and Applied Genetics 38: 226-231.

44. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature. pp. 832-837.

45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

46. Ferreira ID, Rosario VE, Cravo PV (2006) Real-time quantitative PCR with SYBR Green I detection for estimating copy numbers of nine drug resistance candidate genes in Plasmodium falciparum. Malar J 5: 1.

47. Triglia T, Duraisingh MT, Good RT, Cowman AF (2005) Reticulocyte-binding protein homologue 1 is required for sialic acid-dependent invasion into human erythrocytes by Plasmodium falciparum. Mol Microbiol 55: 162-174.

48. Fidock DA, Wellems TE (1997) Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. Proc Natl Acad Sci U S A 94: 10931-10936.

CHAPTER TWO

Modulation of *PF10_0355* alters *Plasmodium falciparum* response to antimalarial drugs

Daria Van Tyne[1], Alessandro D. Uboldi[2], Anthony N. Hodder[2], Julie Healer[2], Alan F. Cowman[2], Dyann F. Wirth[1]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

[2] Walter and Eliza Hall Institute of Medical Research, Victoria, Australia

CONTRIBUTIONS

This manuscript was written by Daria Van Tyne, with input from the other co-authors. Supplemental information can be found in Appendix B. PF10_0355 knock-out parasites and the Southern blot shown in Figure 2.1A were generated by Alessandro D. Uboldi. Anti-PF10_0355 rabbit polyclonal antibodies were generated by the Cowman lab. Identification of interacting proteins shown in Appendix B was conducted at the Taplin Mass Spectrometry Facility in the Gygi lab at Harvard Medical School. Recombinant PF10_0355 and PF10_0348 proteins used for Biacore studies in Appendix B were generated by Anthony N. Hodder. All other data was generated by Daria Van Tyne.

Abstract

Malaria's ability to rapidly adapt to new drugs has allowed it to remain one of the most devastating infectious diseases of humans. Understanding and tracking the genetic basis of these adaptations is critical to the success of treatment and other intervention strategies. The novel antimalarial resistance locus *PF10_*0355 was previously associated with parasite response to halofantrine, and functional validation confirmed that overexpression of this gene lowered parasite sensitivity to both halofantrine and the structurally related antimalarials mefloquine and lumefantrine. *PF10_0355* copy number variation (CNV) appeared to affect parasite drug response in both natural parasite isolates as well as laboratory-generated transgenic lines, however the possible contribution of point mutations within *PF10_0355* remained unknown. Here we further characterize the role of *PF10_0355* in mediating antimalarial drug response in *P. falciparum*. Knockout of *PF10_0355* increased parasite sensitivity to halofantrine, mefloquine and lumefantrine, but not to unrelated antimalarials, further suggesting that resistance is mediated by copy number variation at this locus. We focused on a single nucleotide polymorphism (SNP) encoding a C591S mutation in PF10_0355, which had the strongest association with halofantrine response. Allelic replacement experiments revealed an allele-specific effect on drug response in the absence of copy number variation. Finally, growth competition experiments showed that under drug pressure, the more resistant parasite line out-competed the more sensitive line within a small number of generations. Together these experiments demonstrate that modulation of *PF10_0355* affects drug response in the malaria parasite.

Introduction

Malaria drug resistance poses a serious threat to treatment and control efforts [1] [2]. While resistance loci such as *pfcrt, dhfr* and *pfmdr1* are all known to play a role in mediating *Plasmodium falciparum* drug resistance [3], the precise mechanisms of resistance for many antimalarials are poorly understood. Additionally, these well-known loci do not fully explain the range of responses observed in resistant parasites, suggesting that other loci may be involved in mediating parasite drug response [4] [5].

PF10_0355, also called *MSP3.8* and *PfMSPDBL2*, is a novel antimalarial resistance locus recently identified in a genome-wide association study (GWAS) of 50 global parasite isolates using a high-density SNP array [6]. Overexpression of either the sensitive or resistant allele of *PF10_0355* made parasites less sensitive to halofantrine, mefloquine and lumefantrine, but not to structurally unrelated antimalarials. While this work validated *PF10_0355* as a novel antimalarial resistance locus, the gene was originally identified through association between SNPs and drug response and the possible effects of these mutations on parasite drug response remained unclear.

PF10_0355 is a merozoite surface protein containing a Duffy binding-like (DBL) domain as well as a secreted polymorphic antigen associated with merozoites (SPAM) domain [7]. The PF10_0355 DBL domain binds to red blood cells as a dimer, and binding is dependent on specific metal ions [8]. Additionally, the *PF10_0355* gene is highly polymorphic, appears to be under balancing selection and is a likely target of host immunity [9]. Immunofluorescence studies have shown that PF10_0355 is located on the merozoite surface in both schizonts and free merozoites [7] [8], despite lacking a membrane anchor.

Here we further characterize the role of *PF10_0355* in mediating parasite response to antimalarial drugs. We wanted to probe the differential effects of both CNVs and SNPs on drug sensitivity, and we did this through knockout and allelic replacement experiments. We found that both copy number variation and mutations within *PF10_0355* affect parasite drug response. Additionally, we found that moderate changes in drug response measured in the short-term can cause dramatic differences when parasites are competed with each other in the presence of drug.

Materials and Methods

Parasite culture and drug testing

Parasites were maintained under standard culture conditions [10], in RPMI media supplemented with 5% human O^+ serum and 5% Albumax II. Where indicated, parasites were synchronized with 5% D-sorbitol. Drug testing of knockout and allelic replacement lines was performed by measuring parasite incorporation of tritium-labeled hypoxanthine [11]. Other drug data were generated using a high-throughput SYBR Green I-based assay and results have been previously reported [12]. In all cases, IC_{50}s were calculated using GraphPad Prism (San Diego, CA) using a four-parameter, log-logistic nonlinear regression of fluorescence intensity versus $\log_{10}$-transformed drug concentration. Drug assays were performed with at least two technical replicates, and data shown are averaged over at least two biological replicates.

Generation of knockout vectors and parasite lines

PF10_0355 knockout parasites were generated by double crossover recombination in the 3D7 parasite background using the pCC1 vector ([13]; Uboldi et al. manuscript in prep). *PF10_0355* knockout was confirmed by Southern blot and Western blot using rabbit polyclonal antibodies raised against PF10_0355 [8].

Association between PF10_0355 SNPs and halofantrine response

Whole-genome sequencing of 45 culture-adapted parasite lines from Senegal and genome-wide association studies were performed previously [12]. The SNP Pf_10_001435509 (PlasmoDB v5.0 coordinates; Pf_10_001434265 in PlasmoDB v5.5-v7.2; Pf_10_001434268 in PlasmoDB v9.1) encodes a cysteine-to-serine point mutation at position 591 within the SPAM domain of PF10_0355. SNP calls at this position in 3D7, Dd2 and the *P. reichenowi* Oscar strain were verified previously [6].

Allelic replacement parasites

Allelic replacement vectors were generated by cloning the *PF10_0355* gene sequence minus the first 74 nucleotides from Dd2 (derived allele; DR) and Senegal P26.04 (ancestral allele; AC) parasites into the pHH1 vector [14]. Plasmids were transfected into the Dd2 parasite line and stable transfectants were selected with 2.5 nM WR99210 [15]. Transfectants were twice cycled off WR99210 for two weeks each time, and stable integration and loss of episome was confirmed by Southern blot using ClaI and NotI and a probe generated with the following primers: F: 5'-GGG GAA AGC ATA TAA TAA TAC TAT AGA TGC-3'; R: 5'-CTT GGA GGA ACA AGA ACC CCC TTA TTA TCA-3'. Protein expression of HA-tagged PF10_0355 in the Dd2 0355 DR line was measured by Western blotting with monoclonal anti-HA and anti-LDH antibodies using the Li-cor system (Lincoln, NE).

Immunofluorescence was performed by harvesting mixed-stage parasites growing in culture, pelleting the red blood cells and washing with 1xPBS before fixing in rotating suspension using 4% paraformaldehyde/0.0075% glutaraldehyde for 30 minutes at room temperature. After washing, parasites were permeabilized in rotating suspension using 0.1% Triton X-100 and 3% bovine serum albumin in 1xPBS for 45 minutes at room temperature.

Parasites were then blocked in 3% bovine serum albumin in 1xPBS for at least one hour, followed by rat anti-HA 3F10 antibody (150ng/uL) staining overnight at 4 degrees. After washing, parasites were stained with anti-rat alexa 488 antibodies (1:750 dilution) for 1 hour at room temperature. Cells were mounted using Vectashield Mounting Medium with DAPI (Vector Laboratories, Burlingame, CA) and were imaged using the 100x objective on a Nikon Eclipse TE300 microscope. Images were obtained using MetaMorph software v7.5 (Sunnyvale, CA) with a Hamamatsu C4742-95 camera, and were processed using Adobe Photoshop CS 5.0 (San Jose, CA).

Growth competition experiments

Competition assays were performed by mixing synchronized ring-stage cultures of 0355 DR and 0355 AC lines in equal ratios either in the absence of drug or in the presence of 1nM halofantrine. Parasites were harvested after 1, 4, 8, or 14 asexual cycles and genomic DNA was extracted. Ratios of each line were determined by quantitative real-time PCR (qPCR) using a custom TaqMan assay designed to detect the two different alleles of SNP Pf_10_001435509. Assay primer sequences were: F: 5'-GGG TCA TCA TCT CTT GAA CAA CAC T-3' and R: 5'-TCG CTT TCA TTA GCT ATC TGT TCA ATA TCC-3' and probe sequences were: Allele 1 (0355 AC; VIC): 5'-CAA TTC TAA AGC ACT TCC CTT-3' and Allele 2 (0355 DR; FAM): 5'-ATT CTA AAG CAC ATC CCT T-3'. Ct values were normalized to a standard curve of known genomic DNA mixtures of the two parasite lines.

Results

Knockout of *PF10_0355* makes parasites more sensitive to halofantrine, mefloquine and lumefantrine

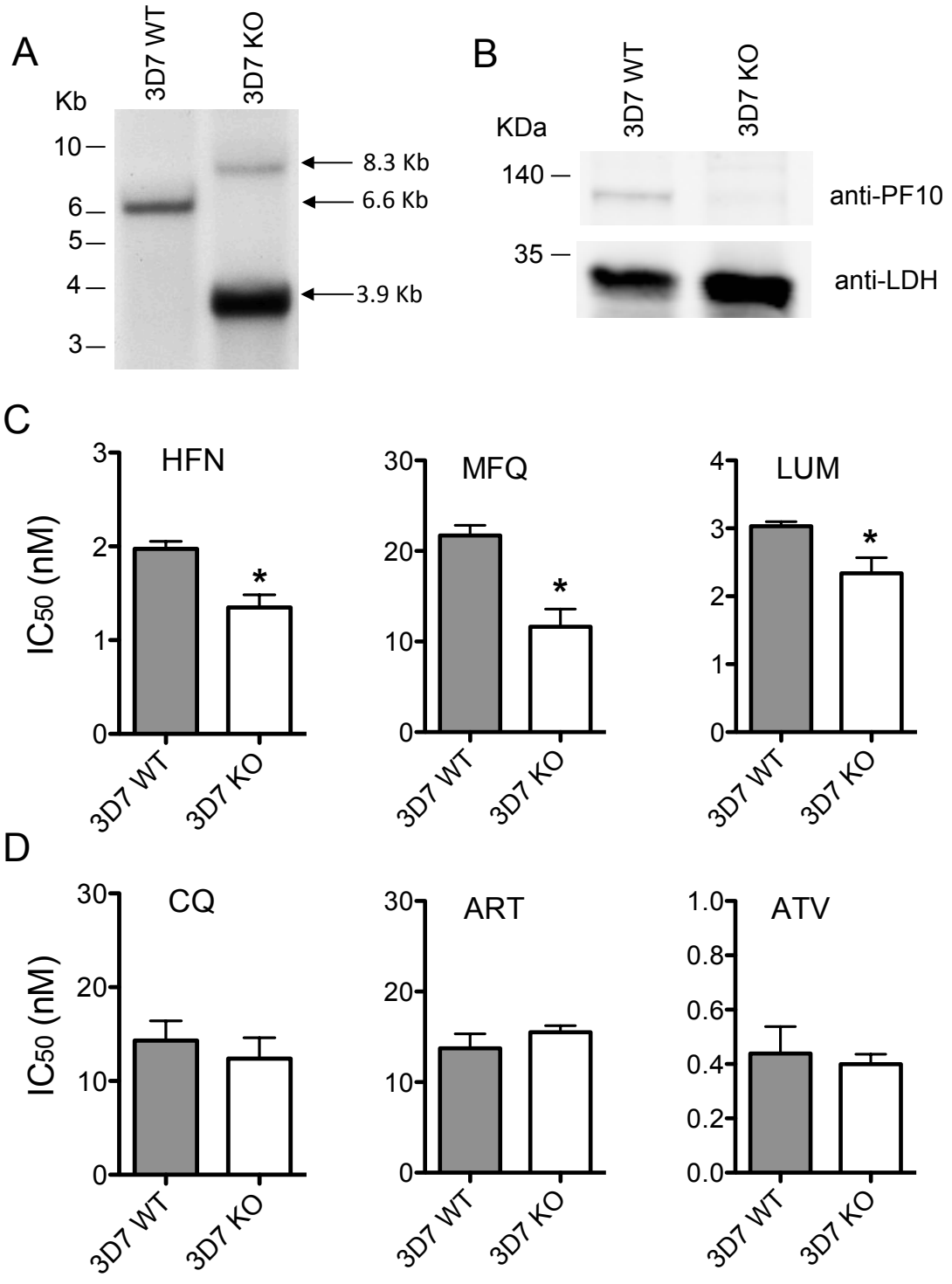
The *PF10_0355* locus was successfully disrupted in the 3D7 parasite line (Figure 2.1). Southern blotting revealed disruption of the endogenous locus (Figure 2.1A; Uboldi et al. manuscript in prep) and Western blotting of schizont-stage parasites using anti-PF10_0355 polyclonal antibodies revealed an absence of protein in the knockout (KO) line (Figure 2.1B). Drug testing by incorporation of tritium-labeled hypoxanthine showed that *PF10_0355* KO parasites were more sensitive than 3D7 wild-type parasites to halofantrine, mefloquine and lumefantrine (Figure 2.1C), but had similar responses to chloroquine, artemisinin and atovaquone (Figure 2.1D).

A mutation within the SPAM domain of *PF10_0355* is associated with parasite response to halofantrine, mefloquine and lumefantrine

SNP 1434268 (PlasmoDB v9.1; 1434265 in PlasmoDB v5.5-v7.2 and 1435509 in PlasmoDB v5.0) on chromosome 10 encodes a non-synonymous, cysteine-to-serine mutation at position 591 within the SPAM domain of *PF10_0355*. This SNP was the non-synonymous SNP within *PF10_0355* most strongly associated with halofantrine response in a recent sequence-based genome-wide association study (GWAS) of 45 Senegalese parasites [12]. Both the *Plasmodium reichenowi* Oscar strain and the *P. falciparum* 3D7 line have cysteine at position 591 [6], suggesting that cysteine is the ancestral (AC) allele and serine is the derived (DR) allele. Only 9 of the 45 Senegalese parasites tested have cysteine at this position and 36 parasites have

Figure 2.1. Knockout of *PF10_0355* increases parasite sensitivity to halofantrine, mefloquine and lumefantrine but not to structurally unrelated drugs. A) Southern blot showing the *PF10_0355* locus in 3D7 wild-type (WT) and knockout (KO) parasites. Digestion of the WT *PF10_0355* locus produces a 6.6Kb band while digestion of the integrated locus produces two bands of 3.9Kb and 8.3Kb. B) Western blot of schizonts from 3D7 WT and KO parasite lines, using polyclonal anti-PF10_0355 and monoclonal anti-LDH antibodies. C) Drug responses in 3D7 WT and KO lines to halofantrine (HFN), mefloquine (MFQ), and lumefantrine (LUM), measured by tritium-labeled hypoxanthine incorporation. D) Drug responses in 3D7 WT and KO lines to chloroquine (CQ), artemisinin (ART), and atovaquone (ATV). Drug responses are shown as mean ± s.e.m. of at least two biological replicates. * $P< 0.05$ by two-tailed, unpaired t-test.

Figure 2.1 (Continued)



serine, resulting in a derived allele frequency of 80% in this population. Parasites with C591 were more sensitive than those with S591 to halofantrine as well as mefloquine and lumefantrine (Figure 2.2, $P < 0.05$ for each drug). The same mutation was not significantly associated with parasite response to chloroquine, artemisinin or atovaquone (data not shown).

To further investigate the association between the C591S mutation and parasite drug response, we generated Dd2 parasite lines where a 3' portion of the endogenous *PF10_0355* locus was replaced by a hemagglutanin (HA)-tagged version of either the derived (0355 DR) or ancestral (0355 AC) allele. Stable integrants containing the 0355 DR or 0355 AC constructs were confirmed by Southern blotting (Figure 2.3A). Western blotting of synchronized lysates from the 0355 DR line using anti-HA antibodies showed proper expression of the tagged protein in schizonts (Figure 2.3B), and immunofluorescence showed correct localization of the HA-tagged protein on the merozoite surface (Figure 2.3C). Drug testing by incorporation of tritium-labeled hypoxanthine revealed that 0355 AR parasites were more sensitive than 0355 DR parasites to halofantrine, mefloquine and lumefantrine (Figure 2.3D), but showed similar responses to chloroquine, artemisinin and atovaquone (Figure 2.3E).

Parasites expressing the *PF10_0355* derived allele out-compete parasites expressing the ancestral allele under drug pressure

We wondered if the differences we observed between 0355 DR and 0355 AC drug responses were relevant when parasites are directly competed with one another. 0355 DR and 0355 AC parasite lines were synchronized and ring-stage parasites were mixed in a 1:1 ratio. Parasite mixtures were allowed to grow in normal media containing no drug or in media

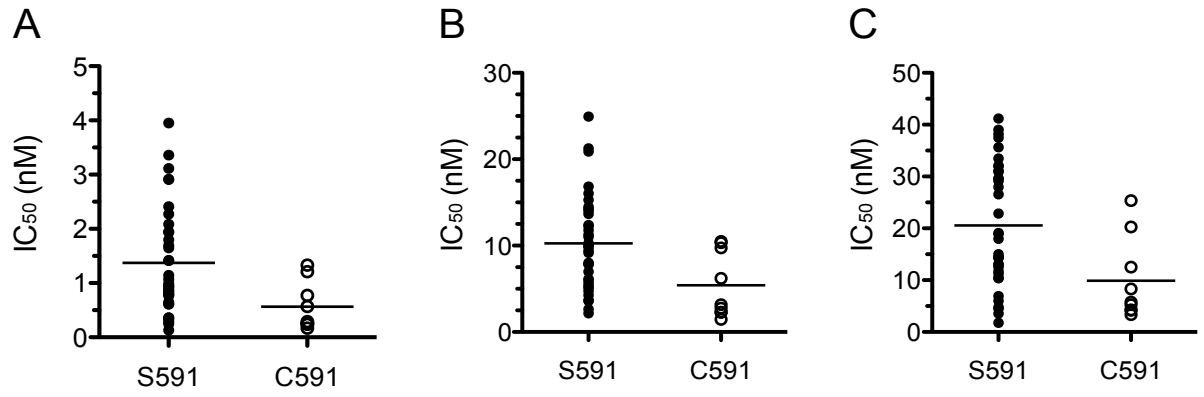
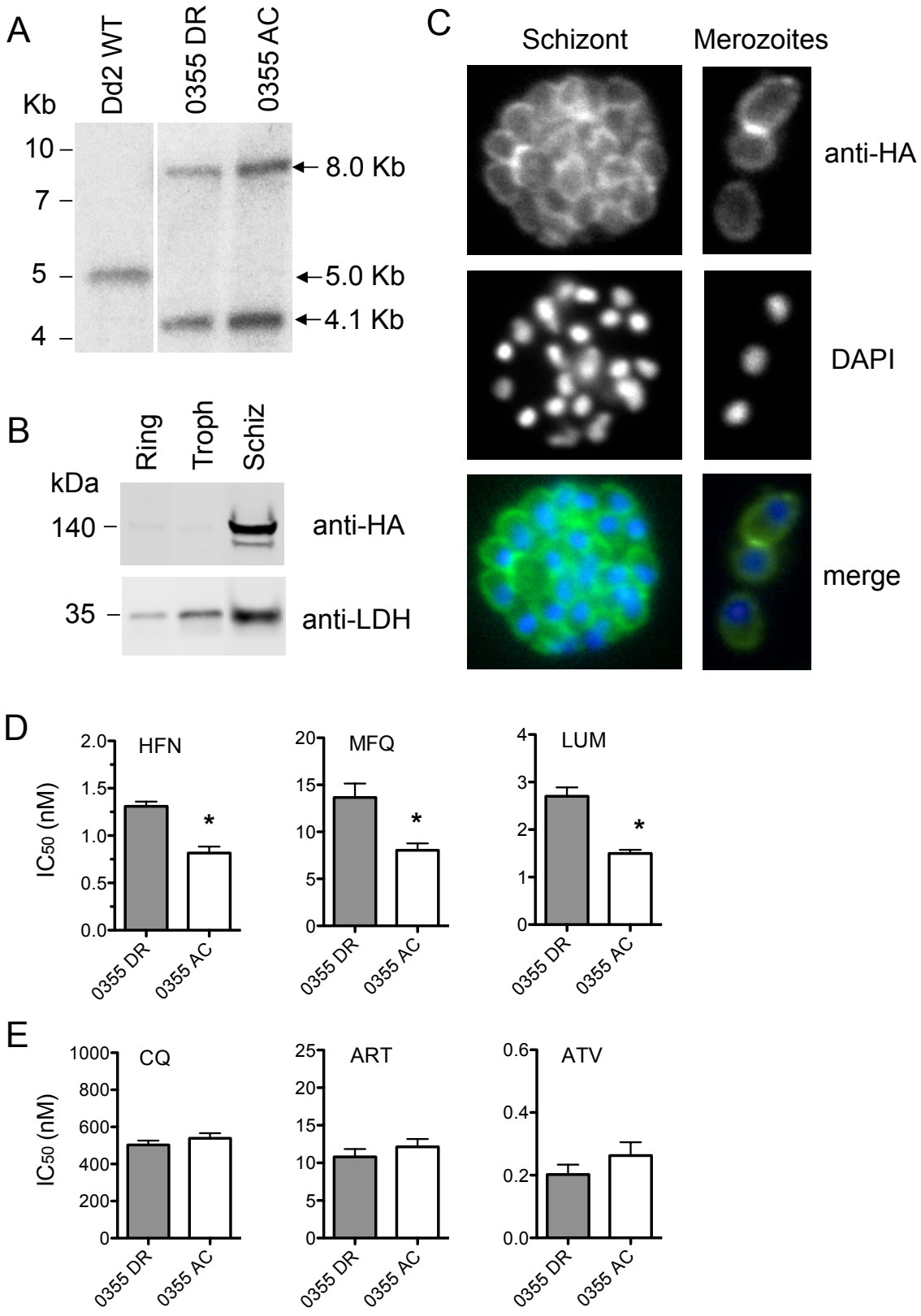


Figure 2.2. A mutation at position 591 within the SPAM domain of *PF10_0355* is associated with parasite response to A) halofantrine, B) mefloquine, and C) lumefantrine in 45 Senegalese parasites. SNP calls and drug data are from [12]; horizontal lines show the mean IC_{50} for each group. *P*-values for two-tailed, unpaired t-test for S591 vs. C591 are $P = .021$ for halofantrine, $P = .016$ for mefloquine and $P = .014$ for lumefantrine.

Figure 2.3. Mutations in the absence of copy number variation affect parasite response to halofantrine and structurally related antimalarials. A) Southern blots of Dd2 wild-type (WT) and PF10_0355 allelic replacement parasites, where the endogenous locus has been replaced with HA-tagged versions of the derived allele (0355 DR, containing S591 from Dd2) or the ancestral allele (0355 AC, containing C591 from Senegal P26.04). Digestion of the WT *PF10_0355* locus with ClaI and NotI generates a band at 5.0Kb while digestion of the integrated locus generates two bands of 4.1Kb and 8.0Kb. B) Western blot of ring, trophozoite (Troph) and schizont (Schiz) stage cultures of the 0355 DR line using anti-HA and anti-LDH antibodies, showing expression of HA-tagged PF10_0355 in schizonts. C) Immunofluorescence of a representative schizont and representative merozoites of the 0355 DR line using anti-HA antibodies and DAPI staining. D) Drug responses in 0355 DR and 0355 AC lines to halofantrine (HFN), mefloquine (MFQ) and lumefantrine (LUM), measured by tritium-labeled hypoxanthine incorporation. E) Drug responses in 0355 DR and 0355 AC lines to chloroquine (CQ), artemisinin (ART) and atovaquone (ATV). Drug responses are shown as mean ± s.e.m. of at least three biological replicates. * $P< 0.05$ by two-tailed, unpaired t-test.

Figure 2.3 (Continued)



containing 1nM halofantrine, and the ratio between 0355 DR and 0355 AC parasites was determined after 1, 4, 8 and 14 asexual replication cycles using a quantitative real-time PCR-based TaqMan SNP assay (Figure 2.4). The ratio of the two parasite lines was unchanged in the absence of drug pressure, remaining close to the 1:1 starting ratio. However when parasites were incubated with 1nM halofantrine, 0355 DR parasites quickly out-competed 0355 AC parasites, reaching a roughly four-fold excess after 8 replication cycles. These results indicate that modest differences in IC_{50}s measured using the standard 72-hour drug assay can be compounded over time to generate large fitness differences when parasites are competed under drug pressure.

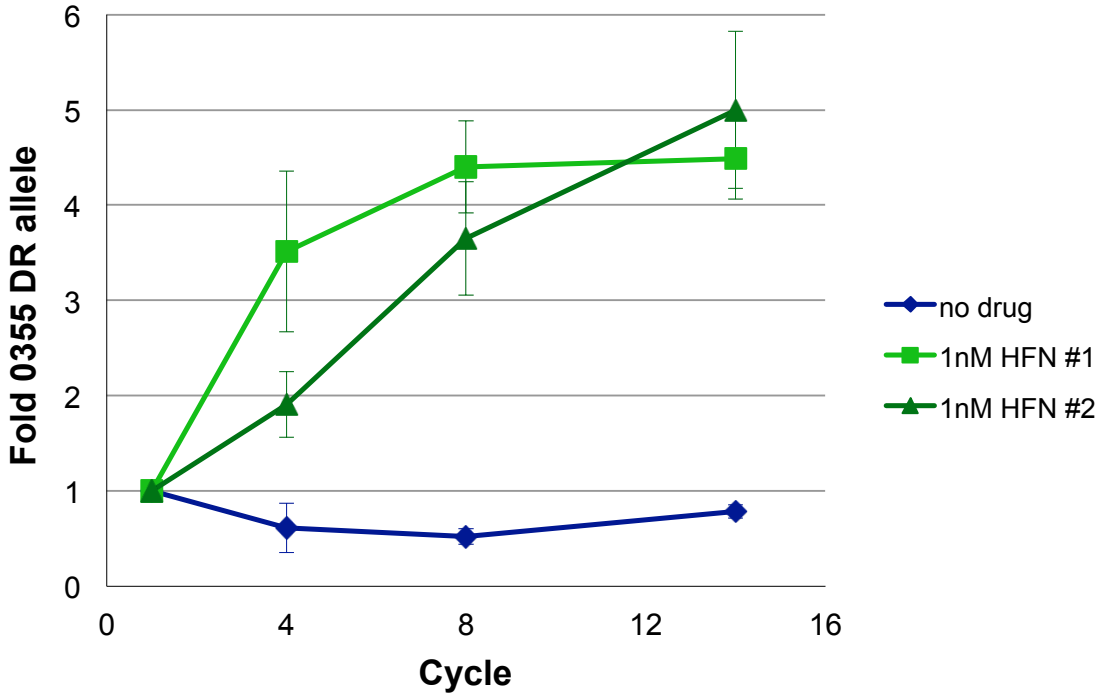


Figure 2.4. 0355 DR parasites outcompete 0355 AC parasites in the presence of halofantrine. Parasites were mixed at equal ratios and grown in no drug or in 1nM halofantrine (HFN). Ratio between 0355 DR and 0355 AC parasites was determined after 1, 4, 8 or 14 asexual cycles by quantitative real-time PCR using TaqMan probes and is shown as fold abundance of the derived allele. Two biological replicates (#1 and #2) are shown and error bars show standard deviation between technical replicates.

Discussion

Drug resistance poses a serious threat to efforts aimed at eliminating malaria from endemic regions, and much attention is currently focused on understanding the molecular basis of resistance. Nonetheless, precise mechanisms of antimalarial drug action and drug resistance remain poorly defined, even for currently used first-line treatments [16]. In the present study, we wanted to know how CNVs and SNPs within *PF10_0355* alter parasite susceptibility to antimalarials, in order to better understand the role of this locus in mediating *P. falciparum* drug response.

Previous work showed that increased *PF10_0355* copy number made parasites more resistant to halofantrine and structurally related antimalarials [6]. Consistent with this result, we found that *PF10_0355* knockout parasites were more sensitive to halofantrine, mefloquine and lumefantrine but not to structurally unrelated drugs. Knockout parasites showed no growth defects compared to wild-type parasites and their increased sensitivity was limited to a specific class of molecules. These findings further support a role for *PF10_0355* copy number variation in modulating parasite drug response, as drug resistance seems to scale with *PF10_0355* copy number in a similar manner to CNVs at other drug resistance loci such as *pfmdr1* and *pfgchI* [17] [18].

We focused on a specific mutation (C591S) within the SPAM domain of *PF10_0355* for two reasons. First, in the original, global GWAS that identified *PF10_0355* as a novel antimalarial resistance locus, the genome wide-significant association signal mapped to the 3' portion of the gene and spanned the C591S mutation [6]. Second, C591S was the mutation within *PF10_0355* that was most strongly associated with parasite response to halofantrine in a recent whole genome sequence-based GWAS [12], though the association did not reach genome-

wide significance. Importantly, only 20% of the 45 sequenced Senegalese parasites retained the ancestral allele at this position, suggesting that selection of some kind has driven the derived allele to high prevalence. Because we used genomic DNA from natural parasite isolates to generate the allelic replacement constructs used here, the 0355 AC parasite line contains additional mutations within *PF10_0355* besides C591S. While it remains possible that mutations other than or in addition to C591S could underlie the changes in drug response that we observed, these results confirm that mutations within *PF10_0355*, in the absence of copy number variation, also mediate parasite drug response.

The precise mechanism of how *PF10_0355* affects antimalarial drug response remains unknown. The PF10_0355 protein localizes to the merozoite surface, despite lacking a membrane anchor. Other merozoite surface proteins are known to form complexes [19], and we found PF10_0355 to be associated with domains of MSP1 by immunoprecipitation (Appendix B, Figure B1). We also wondered whether PF10_0355 binds to halofantrine directly, however attempts to investigate this binding were confounded by high rates of non-specific drug binding to a reference protein (Appendix B, Figure B2). PF10_0355 was recently shown to bind to erythrocytes [8], suggesting that the protein may interact with a specific receptor on the red blood cell surface. Perhaps parasites are better able to bind and invade red blood cells in the presence of specific antimalarials when they have more copies of PF10_0355, or when the protein is in a specific conformation.

Overall, these experiments demonstrate that genetic manipulation of *PF10_0355* affects drug response in *P. falciparum*. Similar to other malaria drug resistance genes, we find effects of both CNVs and SNPs in modulating parasite drug response. Importantly, our findings demonstrate that large differences in parasite fitness can be revealed when parasites are competed over

multiple generations under drug pressure, as they likely would in their natural setting of an infected patient.

Acknowledgments

We thank Bradley Coleman, Manoj Duraisingh, Jeffrey Dvorin and Ulf Ribacke for help designing experiments, interpreting results and helpful discussion. We also thank Daniel Park and Pardis Sabeti for help analyzing whole-genome sequencing data. This work was supported by the Bill and Melinda Gates Foundation.

References

1. malERA Consultative Group on Drugs (2011) A research agenda for malaria eradication: drugs. *PLoS medicine* 8, e1000402

2. Sibley, C.H.*, et al.* (2008) A network to monitor antimalarial drug resistance: a plan for moving forward. *Trends in parasitology* 24, 43-48

3. Ekland, E.H. and Fidock, D.A. (2007) Advances in understanding the genetic basis of antimalarial drug resistance. *Current opinion in microbiology* 10, 363-370

4. Ferdig, M.T.*, et al.* (2004) Dissecting the loci of low-level quinine resistance in malaria parasites. *Molecular microbiology* 52, 985-997

5. Kinga Modrzynska, K.*, et al.* (2012) Quantitative genome re-sequencing defines multiple mutations conferring chloroquine resistance in rodent malaria. *BMC genomics* 13, 106

6. Van Tyne, D.*, et al.* (2011) Identification and Functional Validation of the Novel Antimalarial Resistance Locus PF10_0355 in Plasmodium falciparum. *PLoS Gen* 7, e1001383

7. Singh, S.*, et al.* (2009) A conserved multi-gene family induces cross-reactive antibodies effective in defense against Plasmodium falciparum. *PloS one* 4, e5410

8. Hodder, A.N.*, et al.* (2012) Insights into duffy binding-like domains through the crystal structure and function of the merozoite surface protein MSPDBL2 from P. falciparum. *The Journal of biological chemistry*

9. Ochola, L.I.*, et al.* (2010) Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in Plasmodium falciparum. *Molecular biology and evolution* 27, 2344-2351

10. Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture. *Science* 193, 673-675

11. Desjardins, R.E.*, et al.* (1979) Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrobial agents and chemotherapy* 16, 710-718

12. Park, D.J.*, et al.* (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*

13. Maier, A.G.*, et al.* (2006) Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in Plasmodium falciparum for targeted gene deletion by double crossover recombination. *Molecular and biochemical parasitology* 150, 118-121

14. Harris, P.K.*, et al.* (2005) Molecular identification of a malaria merozoite surface sheddase. *PLoS pathogens* 1, 241-251

15. Fidock, D.A. and Wellems, T.E. (1997) Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proceedings of the National Academy of Sciences of the United States of America* 94, 10931-10936

16. O'Neill, P.M.*, et al.* (2010) The molecular mechanism of action of artemisinin--the debate continues. *Molecules* 15, 1705-1721

17. Sidhu, A.B.*, et al.* (2006) Decreasing pfmdr1 copy number in plasmodium falciparum malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *The Journal of infectious diseases* 194, 528-535

18. Kidgell, C.*, et al.* (2006) A systematic map of genetic variation in Plasmodium falciparum. *PLoS Path* 2, e57

19. Kadekoppala, M. and Holder, A.A. (2010) Merozoite surface proteins of the malaria parasite: the MSP1 complex and the MSP7 family. *International journal for parasitology* 40, 1155-1161

CHAPTER THREE

An overexpression approach to functionally follow-up novel GWAS hits for drug resistance in *Plasmodium falciparum*

Daria Van Tyne[1], Kate M. Fernandez[1], Daniel J. Park[2], Pardis C. Sabeti[2], Dyann F. Wirth[1]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

[2] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

CONTRIBUTIONS

This manuscript was written by Daria Van Tyne, with input from the other co-authors. GWAS plots were generated by Daniel J. Park, with input from Daria Van Tyne. Constructs and transgenic parasites were generated by Daria Van Tyne. Drug testing of laboratory and transgenic lines was carried out by Kate M. Fernandez and Daria Van Tyne.

Abstract

Efforts to control and eliminate *Plasmodium falciparum* malaria are hampered by the rapid emergence and spread of drug resistant parasites. Furthermore, precise mechanisms of antimalarial drug resistance as well as resistance-causing mutations are largely unknown. Genome-wide association studies (GWASs) take an unbiased approach to identify mutations within the *P. falciparum* genome that are associated with parasite drug resistance, and generate hypotheses that require direct testing in order to validate novel hits. Here we take an overexpression approach to functionally follow-up on hits identified in previous GWASs for antimalarial drug resistance. We tested five genes with non-synonymous mutations that were associated with drug resistance in previous GWASs, as well as one hit identified by other approaches that had a strong, but not genome-wide significant, association signal. We overexpressed sensitive and resistance alleles of each candidate gene in sensitive parasite lines and identified putative false-positive associations that did not validate by this approach. We also validated true associations and found that differences in drug response were greater when the alleles were overexpressed in a field parasite from the same population in which the GWAS was conducted. Finally, we found that drug resistance hits from other studies can be combined with GWAS data to generate testable hypotheses. This approach is a useful way to begin to screen the many hits that have been and will be generated by GWASs for drug resistance determinants in the malaria parasite.

Introduction

Malaria caused by *Plasmodium falciparum* kills hundreds of thousands of people each year [1]. Drug resistant parasites are rampant throughout the world, and pose a serious hurdle to malaria treatment and control efforts. Resistant isolates of *P. falciparum* have been identified for nearly all antimalarial drugs that have ever been used, and resistance also appears to be emerging against artemisinin combination therapies (ACTs), the best available current treatment for malaria [2]. Although antimalarial drug resistance is widespread, little is known about the precise mechanisms of action of many antimalarials or the genetic determinants of parasite drug resistance. A better understanding of how drugs act and how parasites become drug resistant could improve antimalarial drug design, allow for more effective resistance monitoring, and better inform drug use.

The genome-wide association study (GWAS) is a useful tool to identify mutations that might underlie drug resistance in *P. falciparum*. GWASs pair genotype data in the form of single nucleotide polymorphisms (SNPs) with drug phenotype data, and identify SNPs that are enriched among drug resistant parasites. GWASs conducted in *P. falciparum* to date have successfully identified both known and novel drug resistance loci [3] [4] [5] [6], as well as loci that may underlie delayed clearance times in parasites exposed to ACTs [7]. The GWAS is an unbiased method that can identify both known and novel mutations associated with parasite drug response, however novel hits still require validation to verify that they are true hits and not false positive associations. Only one malaria GWAS to date has included functional follow-up of a novel locus associated with drug resistance [4]. Genetic validation in the parasite is both time and labor intensive, as transfection efficiency of parasites is exceedingly low and transgenic *P. falciparum* lines can take many months to generate.

Genomic studies such as GWASs in malaria have several advantages over their counterpart studies in humans. Human GWASs are constrained by genome size as well as linkage disequilibrium, which causes association signals to map to large loci containing multiple genes. The malaria genome is 23 megabases, which is roughly on par with the human exome [8], and decreasing costs have made malaria full-genome sequencing approximately the same cost as genotyping arrays [6]. Short linkage disequilibrium, especially among African isolates of *P. falciparum* [9], means that association signals are likely to map to only one or a few genes. One shortfall of malaria genotyping arrays is that markers on the array are generally unable to flag mutations that are not genotyped, however this is not at issue when the entire genome is sequenced and all possible mutations are typed.

Here we use an overexpression approach to functionally validate six novel GWAS hits for parasite response to amodiaquine, chloroquine and dihydroartemisinin. We identified three putative false positive associations and validated two novel loci associated with dihydroartemisinin response. We also show how GWAS data can be combined with other studies to identify associations that do not reach genome-wide significance but may still be biologically relevant. Overall, this approach constitutes a useful method for functional follow-up of novel drug resistance loci in the malaria parasite.

Materials and Methods

Genome-wide association studies and association tests

Genes for functional follow-up were selected from the following association studies: 50 global parasite isolates genotyped using a high density SNP array and tested against 13 antimalarials [4], and 25 or 45 Senegalese parasites that were whole genome sequenced and tested against 12 antimalarials [6] (Table 3.1). "25 Senegal" refers to an initial subset of the 45 parasites that were included in [6]. Association tests used included the multi-marker haplotype likelihood ratio (HLR) [4], single-marker efficient mixed-model association (EMMA) and multi-marker cross-population extended haplotype homozygosity (XP-EHH) [6]. Two hits (*PFD0595w* and *PF10_0342*) were generated by Meta analysis, combining p-values from both EMMA and XP-EHH tests [10] [11].

Functional follow-up

Five hits were chosen for functional follow-up because they contained non-synonymous SNPs that were near or above genome-wide significance for association with drug response by one or more GWAS tests, and because the SNPs were in single exon genes less than 5Kb in length (Figure 3.1). Hits for pyrimethamine were excluded because the overexpression vectors used made parasites pyrimethamine resistant. One additional hit (*PFI0925w*) was chosen because it was identified in a recent study of artemisinin-resistant *P. yoelii* parasites [12], and a mutation within the *P. falciparum* homolog was associated with dihydroartemisinin response by GWAS.

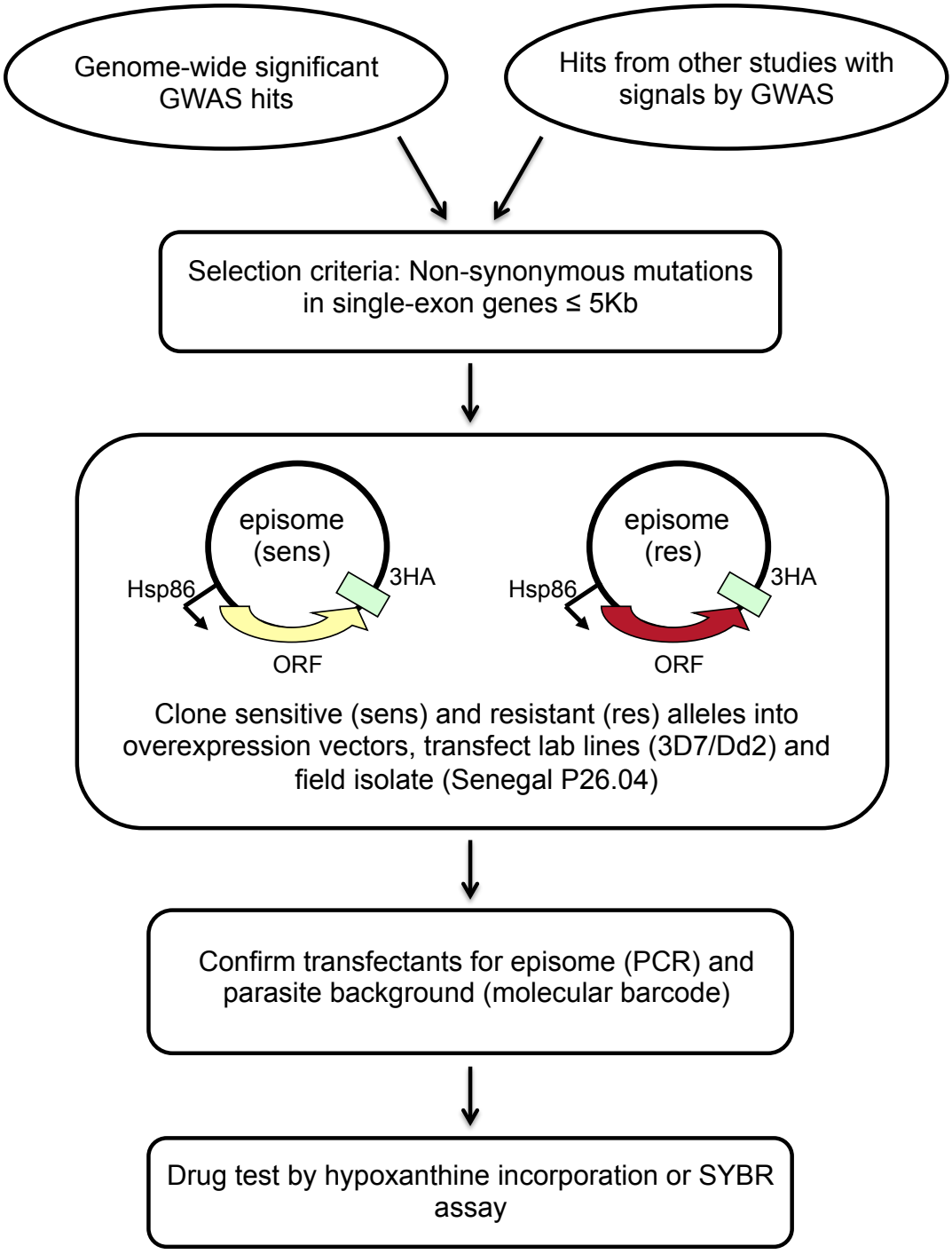
For each gene chosen for functional follow-up, the full-length open reading frame (ORF) minus the stop codon was amplified from two different parasite lines representing the sensitive (sens) and resistant (res) alleles as designated by GWAS (Figure 3.1). ORFs were amplified from the following parasites: *PF11_0302*: Dd2 and Senegal P26.04; *PF10_0344*: Senegal P31.01 and

Table 3.1. Genes associated with antimalarial drug resistance selected for overexpression functional follow-up

Gene ID	Gene Name	Study	Population	Test	SNP	Drug
PF11_0302	Parasitophorous vacuolar protein 1 (Pf-PV1)	Array-based GWAS	50 world-wide	HLR	multiple	Amodiaquine
PF10_0344	Glutamate-rich protein (GLURP)	Seq-based GWAS	25 Senegal	EMMA	A500P	Chloroquine
PFD0595w	Conserved unknown	Seq-based GWAS	25 Senegal	Meta	multiple	Chloroquine
PF10_0342	Probable protein, unknown function	Seq-based GWAS	25 Senegal	Meta	multiple	Dihydro-artemisinin
PFE0355c	Subtilisin-like protease 3 (PfSUB3)	Seq-based GWAS	25 Senegal	EMMA	D397Y	Dihydro-artemisinin
PFI0925w	Gamma-glutamylcysteine synthetase	Other study; Seq-based GWAS	45 Senegal	EMMA	I173K	Dihydro-artemisinin

Figure 3.1. Overview of experimental design. Genes were selected for functional follow-up if they were above or near the cut-off for genome-wide significance by GWAS, or if they were identified by other studies and also had signals by GWAS. Only genes with non-synonymous mutations in single-exon genes less than or equal to 5Kb in size were considered for follow-up. ORFs were cloned from parasite genomic DNA representing the sensitive and resistant alleles for each gene, and overexpression vectors were transfected into laboratory lines (Dd2 or 3D7) and the field isolate Senegal P26.04. Transgenic parasite lines were verified and then tested against a panel of antimalarial drugs. See Methods for additional details.

Figure 3.1 (Continued)



Senegal P26.04; *PFD0595w*: Senegal P31.01 and Senegal Th15.04; *PF10_0342* and *PFE0355c*: Senegal P26.04 and Senegal P19.04; *PFI0925w*: 3D7 and Senegal P51.02. ORFs were cloned into the pBIC009 overexpression vector under the *Hsp86* promoter [4], and were grown in *E. coli* PMC103 bacterial cells.

Parasite culture, transfection and drug testing

Parasites were cultured under standard conditions [13]. Plasmid DNA was isolated from each of the constructs described above and DNA was transfected into Dd2, 3D7 or Senegal P26.04 parasites. Transfectants were selected with 2.5nM WR99210; stable transfectants were confirmed by episome-specific PCR and parasite background was confirmed by SNP-based molecular barcode [14]. Drug testing for the Dd2 parent line and *PF11_0302* transfectants was conducted by measuring incorporation of radiolabeled hypoxanthine [15]. Drug testing for all other transfectants was carried out with a high-throughput, SYBR Green I-based assay [6].

Results and Discussion

Overexpression of GWAS hits has been used previously to validate a novel antimalarial resistance locus in *P. falciparum* [4], and has been listed as a relevant follow-up approach of GWAS hits in human diseases [16] [17]. Genetic manipulation in the malaria parasite is hampered by poor transfection efficiency, long doubling time of the parasite and a small number of available selectable markers. Yet when compared to knockout or allelic replacement experiments, overexpression of genes in *P. falciparum* takes the least amount of time, because once transfected parasites emerge they can be immediately drug tested without having to wait for integration of the construct or parasite sub-cloning. We therefore used overexpression to screen novel drug resistance loci generated by GWAS, in order to determine which hits are likely to be

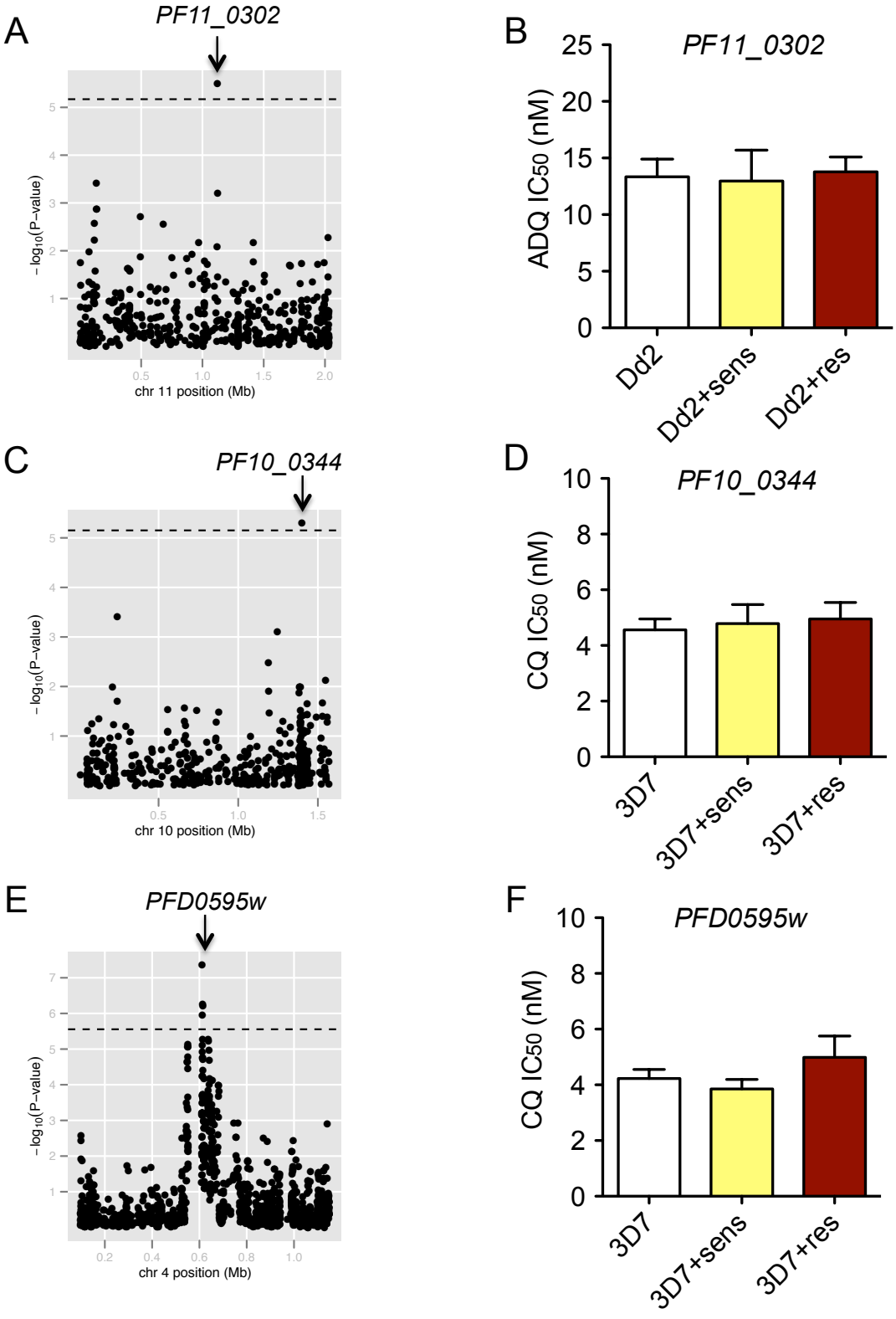
true positive associations that warrant further functional characterization.

Six genes were chosen for overexpression-based validation as novel antimalarial resistance loci (Table 3.1, Figure 3.1). Four hits (*PF11_0302*, *PF10_0344*, *PFD0595w*, and *PF10_0342*) were chosen because they were genome-wide significantly associated with malaria drug response in GWASs conducted previously [4] (Methods). Additionally, *PFE0355c* was the top hit in the genome associated with dihydroartemisinin by the EMMA association test, however the association fell just short of genome-wide significance. Finally, *PFI0925w* was chosen because it was implicated in artemisinin response [12] and was associated with response to dihydroartemisinin by GWAS [6], although the association did not reach genome-wide significance.

We first tried overexpressing the sensitive and resistant alleles for each GWAS hit in the standard laboratory parasite lines Dd2 and 3D7 (Figure 3.2). Three hits (*PF11_0302*, *PF10_0344*, and *PFD0595w*) contained non-synonymous mutations that reached genome-wide significance for association with amodiaquine or chloroquine (Figure 3.2A, C, E), but these hits did not validate when the sensitive and resistant alleles were overexpressed in Dd2 or 3D7 parasite backgrounds (Figure 3.2B, D, F). We therefore consider these hits to be putative false positive associations. Although these hits did not validate by our overexpression approach, it is possible that other follow-up approaches such as allelic replacement, knockout, or overexpression under the endogenous promoter could reveal a role for these genes in modulating parasite response to the drugs they were associated with. The global, SNP array-based GWAS that identified *PF11_0302* included Dd2 as part of the study population, but the GWAS that identified *PF10_0344* and *PFD0595w* included a population of Senegalese parasites that did not

Figure 3.2. Putative false positive GWAS associations. Overexpression follow-up of three GWAS hits suggests they may be false positive associations. A) GWAS plot showing single nucleotide polymorphisms (SNPs) along chromosome 11 and their association with amodiaquine (ADQ) response. Dashed horizontal lines in all GWAS plots indicate the cut-off for genome-wide significant associations, and arrows point to SNPs that map to the gene of interest. B) Summary of ADQ response, as measured by the 50-percent inhibitory concentration (IC_{50}), for Dd2 wild-type parasites (white bar), and Dd2 parasites overexpressing the *PF11_0302* sensitive allele (Dd2+sens, yellow bar) or resistant allele (Dd2+res, red bar). C) GWAS plot showing SNPs along chromosome 10 and their association with chloroquine (CQ) response. D) Summary of CQ IC_{50}s for 3D7 wild-type parasites (white bar), and 3D7 parasites overexpressing the *PF10_0344* sensitive allele (3D7+sens, yellow bar) or resistant allele (3D7+res, red bar). E) GWAS plot showing SNPs along chromosome 4 and their association with CQ response. F) Summary of CQ IC_{50}s for 3D7 wild-type parasites (white bar), and 3D7 parasites overexpressing the *PFD0595w* sensitive allele (3D7+sens, yellow bar) or resistant allele (3D7+res, red bar). Bar charts show mean IC_{50} ± standard deviation of at least three biological replicates for each parasite line.

Figure 3.2 (Continued)

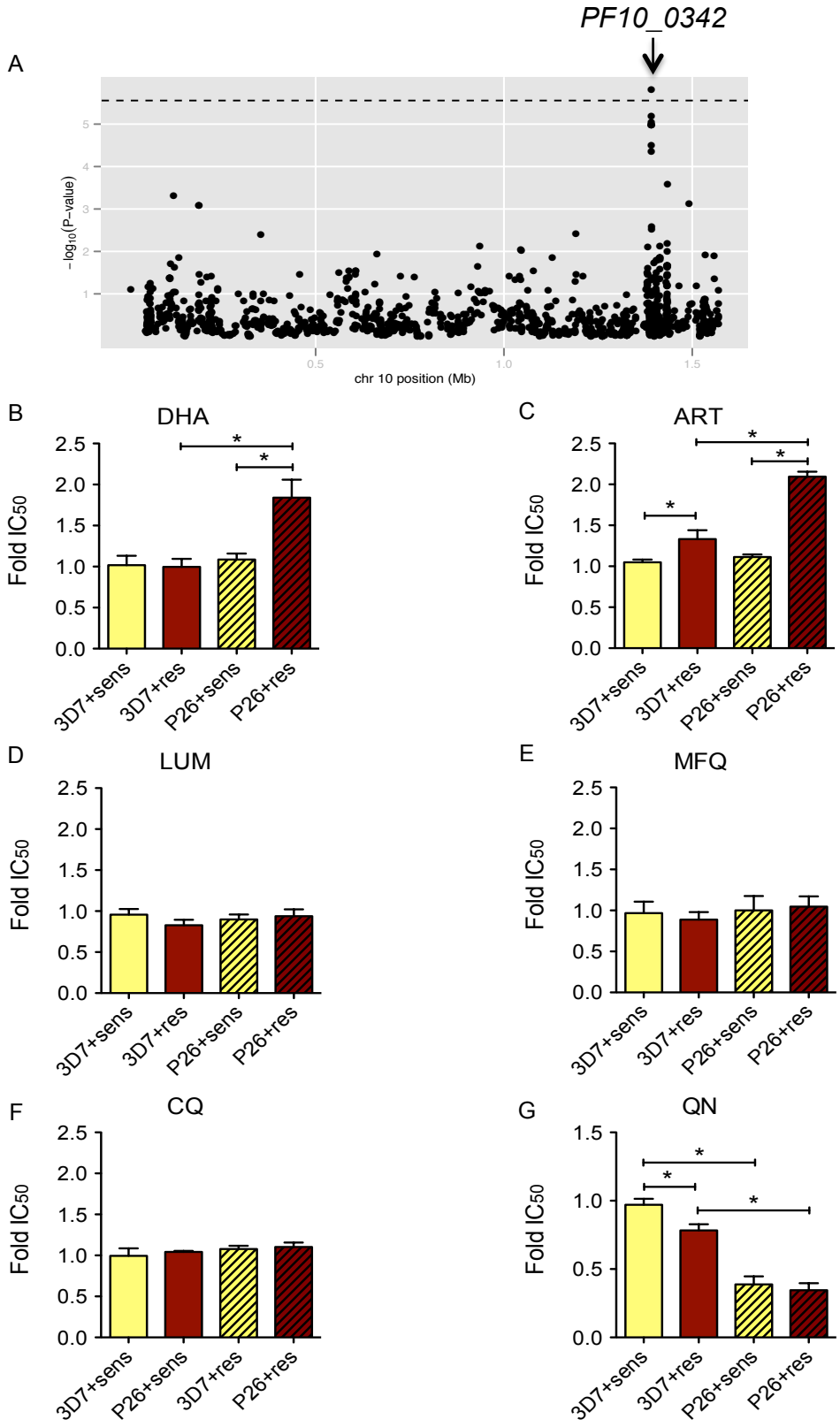


include 3D7. Thus it remains possible that overexpressing these genes in a different parasite background, such as one of the Senegalese parasites used in the GWAS that generated these hits, might reveal an effect for these genes in parasite drug response.

We were able to test the effect of parasite background by overexpressing sensitive and resistant versions of the *PF10_0342* gene in both 3D7 and Senegal P26.04 parasite lines (Figure 3.3). The association signal between mutations in *PF10_0342* and parasite response to dihydroartemisinin (DHA) reached genome-wide significance (Figure 3.3A), but overexpression of the resistant allele in 3D7 parasites had no effect on DHA response (Figure 3.3B). When the sensitive and resistant alleles were overexpressed in the Senegal P26.04 line however, parasites overexpressing the resistant allele were more resistant to DHA than parasites overexpressing the sensitive allele, and these parasites were also more resistant than 3D7 parasites overexpressing the same resistant allele (Figure 3.3B, $P < 0.05$). When parasites were tested against artemisinin (ART), the resistant allele overexpressed in either genetic background made parasites more resistant than overexpression of the sensitive allele, however the effect was greater in the Senegal P26.04 background (Figure 3.3C, $P < 0.05$). None of the overexpressing lines showed differential responses to lumefantrine (LUM, Figure 3.3D), mefloquine (MFQ, Figure 3.3E), or chloroquine (CQ, Figure 3.3F), suggesting that the effect of *PF10_0342* is specific to artemisinin derivatives. Finally, overexpression of the resistant allele in 3D7 made parasites slightly more sensitive to quinine than overexpression of the sensitive allele, however overexpression of either the sensitive or resistant alleles in Senegal P26.04 made parasites much more sensitive (Figure 3.3G, $P < 0.05$). These results suggest that there are other loci that may exert epistatic effects on parasite drug response, and that follow-up experiments may yield different results depending on the parasite background employed.

Figure 3.3. *PF10_0342* is a true positive GWAS association. Functional follow-up of *PF10_0342* shows that overexpression of the resistant allele makes parasites more resistant to dihydroartemisnin (DHA) and artemisinin (ART), when the gene is overexpressed in the field isolate Senegal P26.04. A) GWAS plot showing single nucleotide polymorphisms (SNPs) along chromosome 10 and their association with DHA response. The dashed line indicates the cut-off for genome-wide significance and the arrow points to SNPs that map to *PF10_0342*. B) Summary of DHA response, as measured by fold-change in 50-percent inhibitory concentration (Fold IC_{50}) relative to wild-type parasites, for transgenic 3D7 (solid bars) and Senegal P26.04 (hatched bars) parasite lines overexpressing either the sensitive (+sens, yellow) or resistant (+res, red) alleles of *PF10_0342*. The remaining panels show drug data for the same parasites tested against C) ART, D) lumefantrine (LUM), E) mefloquine (MFQ), F) chloroquine (CQ), and G) quinine (QN). Bar charts show mean Fold IC_{50} ± standard deviation of at least three biological replicates for each parasite line. Bracketed lines show statistical comparisons between parasites, and stars indicate *P*-values < 0.05 by two-tailed, unpaired t-test.

Figure 3.3 (Continued)



We wondered if hits that fell below the cut-off for genome-wide significance might still identify true modulators of parasite drug resistance. The single-marker EMMA GWAS test identified *PFE0355c* as the top signal in the genome for association with DHA response, however the association fell just below the cut-off for genome-wide significance (Figure 3.4A). Similar to *PF10_0342*, overexpression of the resistant allele of *PFE0355c* in the 3D7 background had no effect on DHA response, but an effect was apparent when the resistant allele was overexpressed in Senegal P26.04 parasites (Figure 3.4B, $P < 0.05$). Overexpression of the resistant allele also made both 3D7 and Senegal P26.04 parasites more resistant to ART (Figure 3.4C, $P < 0.05$), but had no effect on LUM, MFQ, CQ or QN (Figure 3.4D, E, F, G). It appears that the threshold for genome-wide significance may be a somewhat arbitrary line and that there can be true association signals that fall below the significance cut-off but are still biologically relevant.

We envision the GWAS as one of many tools that can be used to generate testable hypotheses, and we wondered if GWAS data could be combined with hits from other studies to identify novel drug resistance loci for functional follow-up. *PFI0925w* encodes the *P. falciparum* homolog of gamma-glutamylcysteine synthetase (gamma-GCS), a gene recently found to be down regulated in artemisinin-resistant and chloroquine-resistant *Plasmodium yoelii* lines [12]. We identified a GWAS association between a non-synonymous isoleucine-to-lysine mutation at position 173 within the *P. falciparum* gamma-GCS protein and parasite response to DHA. Although the association fell short of reaching genome-wide significance, it was the strongest DHA association signal within a 100 kilo-base region surrounding *PFI0925w* (Figure 3.5A). We therefore decided to overexpress the sensitive and resistant alleles of *PFI0925w* (as designated by the GWAS signal) in 3D7 parasites. Overexpression of the sensitive allele made parasites

Figure 3.4. *PFE0355c* overexpression alters parasite drug response. The association signal at *PFE0355c* does not reach genome-wide significance but is nonetheless a true positive GWAS hit. A) GWAS plot showing single nucleotide polymorphisms (SNPs) along chromosome 5 and their association with dihydroartemisinin (DHA) response. The dashed line indicates the cut-off for genome-wide significance and the arrow points to a non-synonymous D397Y SNP within *PFE0355c*. B) Summary of DHA response, as measured by fold change in 50-percent inhibitory concentration (Fold IC_{50}) relative to wild-type parasites, for transgenic 3D7 (solid bars) and Senegal P26.04 (hatched bars) parasite lines overexpressing either the sensitive (+sens, yellow) or resistant (+res, red) alleles of *PFE0355c*. Remaining panels show drug data for the same parasites tested against C) artemisinin (ART), D) lumefantrine (LUM), E) mefloquine (MFQ), F) chloroquine (CQ), and G) quinine (QN). Bar charts show mean Fold IC_{50} ± standard deviation of at least three biological replicates for each parasite line. Bracketed lines show statistical comparisons between parasites, and stars indicate *P*-values < 0.05 by two-tailed, unpaired t-test.

Figure 3.4 (Continued)

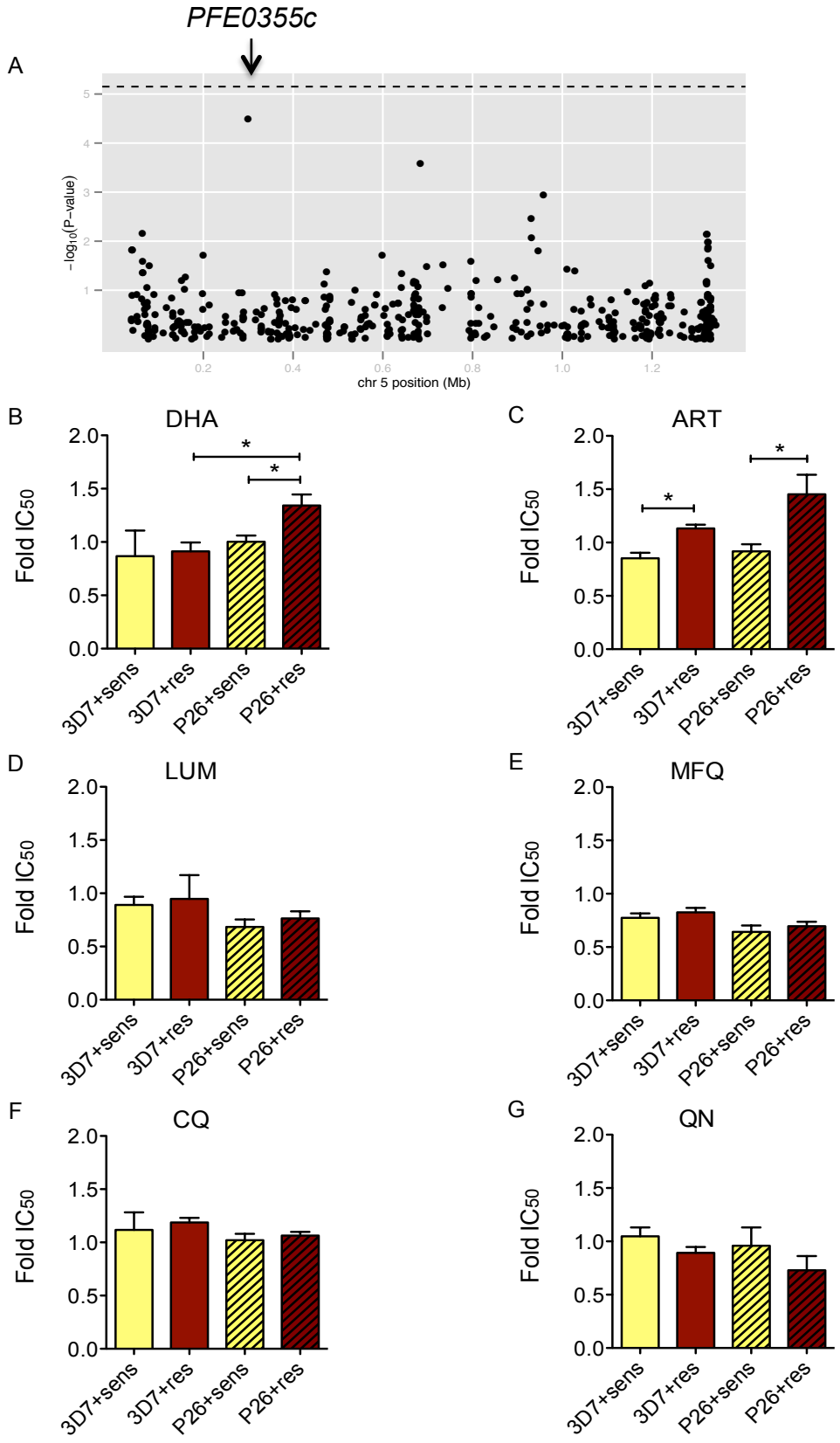
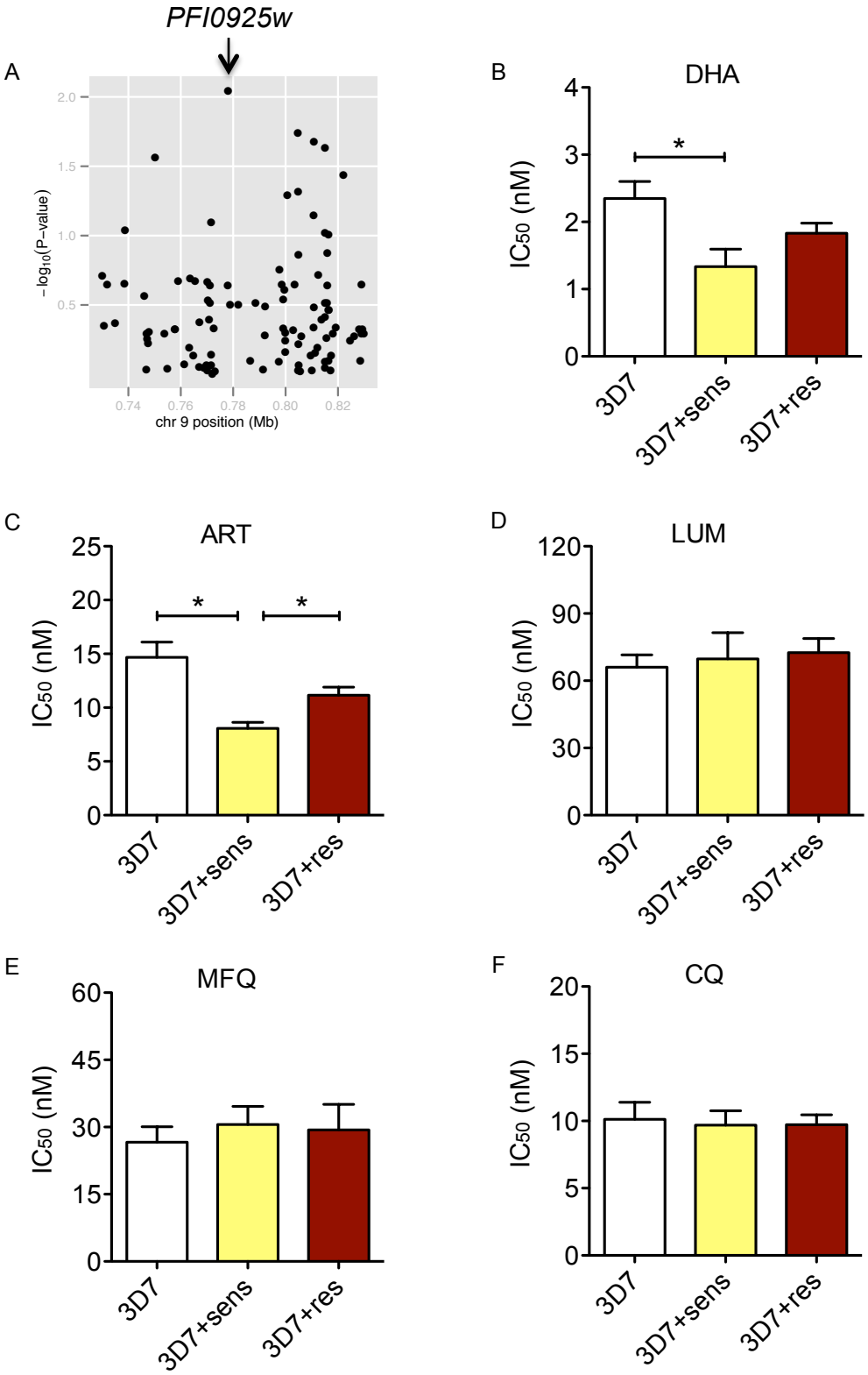


Figure 3.5. Functional follow-up supports a role of *PFI0925w* in parasite response to artemisinin and its derivatives. Genes identified by other approaches can be combined with GWAS data to generate testable hypotheses. A) GWAS plot showing single nucleotide polymorphisms (SNPs) in a 100Kb window of chromosome 9 and their association with dihydroartemisinin (DHA) response. The arrow points to a non-synonymous I173K SNP within *PFI0925w*, which is the strongest association within this genomic region. B) Summary of DHA response, as measured by the 50-percent inhibitory concentration (IC_{50}), for 3D7 wild-type parasites (white bar), and 3D7 parasites overexpressing the *PFI0925w* sensitive allele (3D7+sens, yellow bar) or resistant allele (3D7+res, red bar). Remaining panels show drug data for the same parasite lines tested against C) artemisinin (ART), D) lumefantrine (LUM), E) mefloquine (MFQ), and F) chloroquine (CQ). Bar charts show mean IC_{50} ± standard deviation of at least three biological replicates for each parasite line. Bracketed lines show statistical comparisons between parasites, and stars indicate *P*-values < 0.05 by two-tailed, unpaired t-test.

Figure 3.5 (Continued)



more sensitive to DHA (Figure 3.5B, $P < 0.05$). Parasites overexpressing the sensitive allele were more also sensitive to ART, however overexpression of the resistant allele made parasites slightly more resistant to ART compared to overexpression of the sensitive allele (Figure 3.5C, $P < 0.05$). This suggests that there may be an allele-specific effect on drug response, as indicated by the GWAS signal. Drug response was unchanged when 3D7 wild-type and overexpressing parasites were tested against LUM, MFQ, and CQ (Figure 3.5D, E, F). Transgenic parasites were not generated in a field isolate for this gene, but doing so may reveal a more dramatic or differential effect on parasite drug response, as it has for *PF10_0342* and *PFE0355c*.

GWASs have the potential to identify multiple loci that individually have a small impact on malaria drug response, but when considered together could make the difference between treatment success and failure. It seems clear that parasite response to artemisinin and its derivatives may involve multiple genes in different biological pathways. A recent analysis of the HB3 x Dd2 genetic cross revealed multiple genetic loci associated with artemisinin response, as well as a range of responses among the progeny that was greater than that of the two parent lines [18]. None of the loci identified overlap with the genes studied here, however the parasites analyzed were derived from different populations and only variants present in either parent could be detected by the genetic cross. Furthermore, none of the current candidate marker genes for artemisinin resistance were associated with reduced *in vivo* sensitivity to artemisinin among parasites on the Thai-Cambodian border [19], suggesting that additional loci are involved in modulating parasite drug response in this area.

This study highlighted three novel genes that appear to be involved in modulating parasite response to artemisinin and its derivatives, but a mechanistic understanding of their roles in parasite drug resistance remains unclear. *PF10_0342* is highly polymorphic and shows peak

transcription during schizont and merozoite stages [20], and was also identified as a possible novel erythrocyte invasion and human virulence gene [21]. *PFE0355c* is a putative serine protease and little is known beyond its expression and localization within the parasite and parasitophorous vacuole [22] [23]. *PFI0925w* is the rate-limiting enzyme in the synthesis of glutathione, and this pathway is believed to play a role in mitigating oxidant stress on the parasite during infection. Unsuccessful attempts to knockout *PFI0925w* suggest that the gene is essential [24], however knockout of the gene in *P. berghei* showed that gamma-GCS is dispensable in that parasite, but that it does play a significant role in mosquito-stage parasite development [25]. Our finding that parasites overexpressing *PFI0925w* were more sensitive to ART and DHA is in agreement with previous results, namely that more resistant parasites show lower expression of the gene [12]. Nonetheless, it remains unclear exactly how *PFI0925w* mediates parasite response to these drugs.

Overall, these findings show that overexpression is a worthwhile and useful method for functional follow-up of GWAS hits in the malaria parasite. We found that half of the loci we studied, which were all genome-wide significantly associated with parasite drug resistance, showed no effect on parasite drug response when tested by our approach. We also found that overexpression of two novel genes associated with DHA response did change parasite response to artemisinin and its derivatives, and that the effect was strongest when the genes were overexpressed in a field isolate rather than a laboratory parasite line. Finally, we show how GWAS data can complement other studies of genetic loci that mediate parasite drug response, and our findings suggest that combining independent approaches can identify additional novel drug resistance loci in the malaria parasite.

Acknowledgments

The authors would like to thank Sarah Volkman, Allison Demas, Jon Herman and Ulf Ribacke for useful discussion. Funding for this work was provided by the Bill and Melinda Gates Foundation.

References

1. WHO (2011) World Malaria Report 2011

2. Dondorp, A.M., *et al*. (2009) Artemisinin resistance in Plasmodium falciparum malaria. *NEJM* 361, 455-467

3. Mu, J., *et al*. (2010) Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nature genetics* 42, 268-271

4. Van Tyne, D., *et al*. (2011) Identification and Functional Validation of the Novel Antimalarial Resistance Locus PF10_0355 in Plasmodium falciparum. *PLoS Gen* 7, e1001383

5. Yuan, J., *et al*. (2011) Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* 333, 724-729

6. Park, D.J., *et al*. (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*

7. Cheeseman, I.H., *et al*. (2012) A major genome region underlying artemisinin resistance in malaria. *Science* 336, 79-82

8. Ng, S.B., *et al*. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272-276

9. Manske, M., *et al*. (2012) Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* 487, 375-379

10. Whitlock, M.C. (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of evolutionary biology* 18, 1368-1373

11. Zaykin, D.V. (2011) Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *Journal of evolutionary biology* 24, 1836-1841

12. Witkowski, B., *et al*. (2012) Evidence for the contribution of the hemozoin synthesis pathway of the murine Plasmodium yoelii to the resistance to artemisinin-related drugs. *PloS one* 7, e32620

13. Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture. *Science* 193, 673-675

14. Daniels, R., *et al*. (2008) A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. *Malaria journal* 7, 223

15. Desjardins, R.E., *et al*. (1979) Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrobial agents and chemotherapy* 16, 710-718

16. Frayling, T.M. and McCarthy, M.I. (2007) Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here? *Diabetologia* 50, 2229-2233

17. Freedman, M.L., *et al*. (2011) Principles for the post-GWAS functional characterization of cancer risk loci. *Nature genetics* 43, 513-518

18. Beez, D., *et al*. (2011) Genetic predisposition favors the acquisition of stable artemisinin resistance in malaria parasites. *Antimicrobial agents and chemotherapy* 55, 50-55

19. Imwong, M., *et al*. (2010) Exploring the contribution of candidate genes to artemisinin resistance in Plasmodium falciparum. *Antimicrobial agents and chemotherapy* 54, 2886-2892

20. Ochola, L.I., *et al*. (2010) Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in Plasmodium falciparum. *Molecular biology and evolution* 27, 2344-2351

21. Frech, C. and Chen, N. (2011) Genome comparison of human and non-human malaria parasites reveals species subset-specific genes potentially linked to human disease. *PLoS computational biology* 7, e1002320

22. Withers-Martinez, C., *et al*. (2004) Subtilisin-like proteases of the malaria parasite. *Molecular microbiology* 53, 55-63

23. Sargeant, T.J., *et al*. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome biology* 7, R12

24. Patzewitz, E.M., *et al*. (2012) Dissecting the role of glutathione biosynthesis in Plasmodium falciparum. *Molecular microbiology* 83, 304-318

25. Vega-Rodriguez, J., *et al*. (2009) The glutathione biosynthetic pathway of Plasmodium is essential for mosquito transmission. *PLoS pathogens* 5, e1000302

CHAPTER FOUR

Monitoring antimalarial drug response in *Plasmodium falciparum* field isolates using an *ex vivo* DAPI assay

Daria Van Tyne[1], Papa Diogoye Sène[2]*, Clarissa Valim[1,3]*, Baba Dieye[2], Rachel F. Daniels[1], Amanda K. Lukens[1,3], Mouhamadou Ndiaye[2], Amy K. Bei[1], Yaye Die Ndiaye[2], Omar Ndir[2], Souleymane Mboup[4], Sarah K. Volkman[1,3,5], Dyann F. Wirth[1,3], Daouda Ndiaye[2]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston 02115, USA

[2] Laboratory of Parasitology and Mycology, Université Cheikh Anta Diop, Dakar BP5005, Senegal

[3] Infectious Disease Initiative, Broad Institute, Cambridge 02141, USA

[4] Laboratory of Bacteriology and Virology, Université Cheikh Anta Diop, Dakar BP5005, Senegal

[5] College of Nursing and Health Sciences, Simmons College, Boston 02115, USA

* These authors contributed equally to this work

CONTRIBUTIONS

This manuscript was written by Daria Van Tyne, with input from the other co-authors. Supplemental information can be found in Appendix C. Parasite samples were collected by Daouda Ndiaye, along with Senegalese physicians and researchers overseen by him. *Ex vivo* drug data was collected by Daria Van Tyne, Papa Diogoye Sène, Baba Dieye, Mouhamadou Ndiaye, Yaye Die Ndiaye and Amy Bei. Drug data was analyzed by Daria Van Tyne, Papa Diogoye Sène and Daouda Ndiaye. HRM genotyping was performed by Rachel Daniels. *In vitro* drug testing was performed by Amanda Lukens, Justin Becker and Meaghan Galligan. Statistical analyses were performed by Clarissa Valim and Daria Van Tyne. In Appendix C, Figures C1 and C2 were generated by Daria Van Tyne.

Abstract

BACKGROUND: Malaria treatment efforts are hindered by the rapid emergence and spread of drug resistant parasites. Simple assays to monitor parasite drug response directly from patient samples are needed, as they can detect drug resistance before it becomes clinically apparent and inform changes in treatment policy to prevent the spread of resistant parasites.

METHODS: We surveyed *P. falciparum* malaria cases in a clinic in Thiés, Senegal from 2008-2010 and used a DAPI-based drug assay to test parasite response to amodiaquine, chloroquine and mefloquine in approximately 300 clinical isolates gathered directly from patients (*ex vivo*). We genotyped known drug resistance-associated mutations and culture adapted a subset of parasites to compare *ex vivo* and *in vitro* drug responses.

RESULTS: The prevalence of resistance-associated mutations in *pfcrt* and *pfmdr1* changed between 2008 and 2011, with some mutations increasing and others decreasing in prevalence. Furthermore, these mutations were associated with expected changes in *ex vivo* drug response. We also observed strong concordance between *ex vivo* and *in vitro* 50-percent inhibitory concentrations (IC_{50}s) in a subset of parasite samples that were culture adapted.

CONCLUSIONS: Directly measuring *ex vivo* parasite drug response complements resistance mutation genotyping, and both approaches provide important information about parasite drug response in field samples.

Introduction

Plasmodium falciparum malaria has an enormous public health impact, infecting millions and killing hundreds of thousands of people each year [1]. Drug resistance further magnifies the burden of this disease, as the malaria parasite has evolved resistance to nearly every antimalarial drug developed to date. Recent reports of parasites with reduced susceptibility to the most effective current treatment, artemisinin combination therapies (ACTs) [2] [3], underscore the importance of closely monitoring parasite drug response and optimizing control strategies to prevent the spread of resistant parasites.

Routine monitoring of known drug resistance markers can determine whether antimalarials will be effective against the parasites circulating in an endemic area, and can signal the reemergence of drug-sensitive parasites after a drug is removed from the treatment arsenal. Mutations in a number of loci have been shown to contribute to antimalarial drug resistance, including *pfcrt* and *pfmdr1*, among others [4]. These markers are important surveillance tools for malaria-endemic areas, and can be paired with clinical and parasite drug response data to better inform drug use and extend the therapeutic life of current and future treatments [5].

Assays that directly measure antimalarial drug response in naturally circulating parasite populations can detect decreases in drug efficacy before resistance becomes clinically evident and widespread. Two classic tests for parasite drug resistance measure inhibition of parasite growth by microscopic measurement of schizont maturation [6] or parasite incorporation of the radiolabeled metabolite hypoxanthine [7]. More recently, assays measuring parasite drug response by DNA staining with DAPI or SYBR Green I have been developed [8] [9]. DNA stain-based assays are more amenable to measuring parasite drug response directly from patient samples (*ex vivo*) because they do not generate radioactive waste and do not rely on microscopy

for data acquisition, which can be variable and reader-dependent. Malaria drug assays have also been successfully deployed in field settings, including DNA stain-based assays [10] [11] and ELISA-based assays that measure parasite-produced lactate dehydrogenase (LDH) and histidine-rich protein II (HRP2) [12] [13].

Here we tested whether parasites circulating in Thiés, Senegal are becoming more or less resistant to current and previously-administered antimalarial drugs. We directly tested parasite drug response to amodiaquine, chloroquine and mefloquine using a previously developed DAPI-based *ex vivo* assay [11]. We also genotyped known drug resistance-associated mutations in *pfcrt* and *pfmdr1* to see if mutation prevalence changed over time and whether mutations correlate with *ex vivo* parasite drug response. Finally, we tested whether parasite drug responses in our *ex vivo* assay were good surrogates for *in vitro* drug responses following culture adaptation.

Materials and Methods

Study Population

Individuals seeking treatment for uncomplicated *Plasmodium falciparum* malaria at the Section de Lutte Antiparasitaire (SLAP) clinic in Thiés, Senegal during the fall transmission seasons of 2008-2010 were tested for malaria infection by microscopy and rapid diagnostic test (RDT). *P. falciparum*-positive patients were enrolled if they met the eligibility criteria described below and the patient or their legal guardian provided informed written consent or assent. Eligibility criteria were: patients older than 2 years, axillary temperature above 37.5^{o}C or history of fever within the preceding 24 hours, infection with only *P. falciparum*, no recent antimalarial drug use, and a hemoglobin level greater than 6 g/dL. Patients with symptoms of severe malaria were excluded and referred for appropriated care. Study protocols and informed consent

documents were approved by the Institutional Review Boards of both the Senegal Ministry of Health and the Harvard School of Public Health.

Among 682 patients with uncomplicated malaria screened at the SLAP clinic in 2008-2010, 291 patients met the inclusion criteria for this study. The subset of included patients was comparable to the larger set of screened patients with respect to demographic parameters (age, gender) and most clinical characteristics (temperature, hematocrit, weight) with the exception of parasitemia (Table 4.1).

Sample Collection and DAPI *Ex Vivo* Testing

From each subject, 5-10mL venous blood were collected and processed on the same day. Approximately 1mL of blood was spotted onto Whatman FTA^{TM} filter paper cards for subsequent DNA extraction; the remaining blood was spun at 1500rpm for 10 minutes, plasma and buffy coat were removed and infected red blood cells were washed twice with unsupplemented RPMI media. Aliquots of each sample were cryopreserved in Glycerolyte 57 (Fenwal) supplemented with AB^{+} serum for subsequent culture adaptation.

Parasites were drug tested using the previously described DAPI *ex vivo* assay [11]. Briefly, 180µL of supplemented RPMI media containing parasitized erythrocytes at 2% hematocrit were distributed into 96-well plates preloaded with 20µL serial dilutions of amodiaquine (USP 1031004), chloroquine (Sigma 6628) and mefloquine (Sigma M2319). Drug concentrations ranged from less than 1nM to greater than 1µM and each plate included 6-8 negative control wells with media only. Plates contained two wells of each drug concentration and were prepared in a single batch and frozen prior to use. When possible, samples with parasitemia greater than 1% were diluted into leukocyte-free donor O^{+} erythrocytes to a final plating parasitemia of 0.5-1%.

Table 4.1. Clinical Parameters in Screened Patients and the Subset Included in the DAPI *Ex Vivo* Assay

	All screened patients	DAPI patients	*P*
Number	682	291	-
Gender (% male)	66	64	0.68
Age (years)	21 (15, 29)	21 (15, 27)	0.48
Weight (kg)	57 (45, 65)	55 (42, 65)	0.64
Temperature (^{o}C)	38.2 (37.2, 39.9)	38.4 (37.4, 40.0)	0.12
Hematocrit (%)	38 (32, 40)	38 (33, 41)	0.42
Parasitemia (%)	0.50 (0.20, 1.00)	0.65 (0.40, 1.01)	<.0001

Median values (with interquartile ranges) are reported for age, weight, temperature, hematocrit and parasitemia.

P-values were calculated using Pearson X^2 for categorical variables and Wilcox rank-sum test for continuous variables.

Parasites were cultured 48-72 hours at 37^{o}C under standard conditions (1% O_2, 5% CO_2, 94% N_2) before addition of 4',6-diamidino-2-phenylindole (DAPI) solution, as described previously [8] [11]. Data were collected by measuring relative fluorescence units (RFUs) using a Fluoroskan plate reader (Thermo Scientific; ex 358nm, em 461nm). 3D7 parasites were tested on each batch of drug plates to control for batch variation.

DNA Extraction, Clonality of Infection and HRM Genotyping

DNA was extracted from 4-5 circular 6mm punches of blood preserved on Whatman FTA^{TM} filter paper cards using either a QIAmp DNA Blood Mini Kit (Qiagen) or a Maxwell DNA IQ Casework Sample Kit (Promega). Parasite genomic DNA was quantified by quantitative PCR and clonality of infection, defined as monoclonal or polyclonal, was assessed using the 24-SNP molecular barcode [14].

High resolution melt (HRM) technology was used to genotype a set of single nucleotide polymorphisms (SNPs) associated with reduced drug sensitivity [15]. Briefly, 0.01ng of parasite template, as quantified by qPCR [14], was used for each 5μL reaction, which included 2.5x LightScanner Master Mix with LCGreen Plus dsDNA dye (Idaho Technology, Inc.). HRM analysis and genotype determination was performed on a LightScanner-384 (Idaho Technology, Inc).

Culture Adaptation and *In Vitro* Drug Testing

To assess whether *ex vivo* drug responses were reproducible *in vitro*, we culture adapted 16 parasites derived from monoclonal infections collected in 2009. Culturing was conducted under standard conditions [16] with gentle shaking. Parasites were *in vitro* drug tested against a panel of known antimalarials using a SYBR Green I-based drug assay [17] with modifications for 384-well format. Briefly, synchronized ring-stage parasites were cultured with serial dilutions

of test compounds in 40μL supplemented RPMI media at 1% hematocrit and 1% initial parasitemia, in black clear-bottom plates (Greiner Bio-one 781090). Following 72-hr incubation, a 1:5,000 dilution of SYBR Green I dye (Invitrogen S7563) was added and plates were stored at room temperature until fluorescence signal was read on a Spectramax M5 plate reader (Molecular Devices; ex 480nm, em 530nm).

Calculation of IC_{50}s and Data Exclusion

Fluorescence data from drug assays were analyzed using GraphPad Prism (San Diego, CA) through a four-parameter, log-logistic nonlinear regression of fluorescence intensity versus $\log_{10}$-transformed drug concentration. To include control wells with no drug in the analysis, 1nM was added to each concentration value. Dose-response curves were visually inspected for fit of the sigmoidal dose-response model. Among 291 samples tested in the DAPI *ex vivo* assay, 9 samples were considered assay failures due to no parasite growth or assay contamination and were excluded from further analysis. We excluded an additional 2 patient samples with an initial parasitemia below 0.1% and 30 samples with an initial parasitemia above 1.5%, because there was no clear association between initial parasitemia and fluorescence intensity in the no-drug wells, perhaps due to insufficient growth or saturation.

Drug curves that did not exhibit the standard sigmoidal dose-response shape were classified as either fitting an exponential or linear decay model and had their IC_{50}s estimated through these alternative models, or were excluded. Excluded curves were generally flat, possibly due to no parasite growth or assay contamination. When IC_{50}s from technical replicates could not be estimated due to a single outlier point, this point was excluded.

DAPI Assay Validation

Fluorescence intensity in no-drug wells was linearly correlated with input parasitemia (ρ = 0.51, linear slope $P < .0001$) (Figure 4.1A). Fluorescence signal was not substantially affected by human DNA from white blood cells or DNA from dead parasites, as suggested by the weak correlation between initial parasitemia and fluorescence signal in the maximum-drug wells (ρ = 0.33). Dynamic range was assessed through the signal-to-noise ratio (SNR) of each assay, measured by dividing fluorescence signal from no-drug wells by fluorescence signal from maximum drug wells. Median SNR among all samples tested was 4:1 (Interquartile Range = 3:1, 6:1) (Figure 4.1B). SNR was weakly correlated with parasitemia (ρ = 0.26, $P < 0.001$) and was not different across years (Kruskal-Wallis $P = 0.29$). We also calculated the Z'-factor for each assay using the following equation: Z' = 1- [(3 standard deviations of positive controls + 3 standard deviations of negative controls)/absolute difference between negative and positive controls] [18]. The median Z'-factor among all assays was 0.61 (Figure 4.1C).

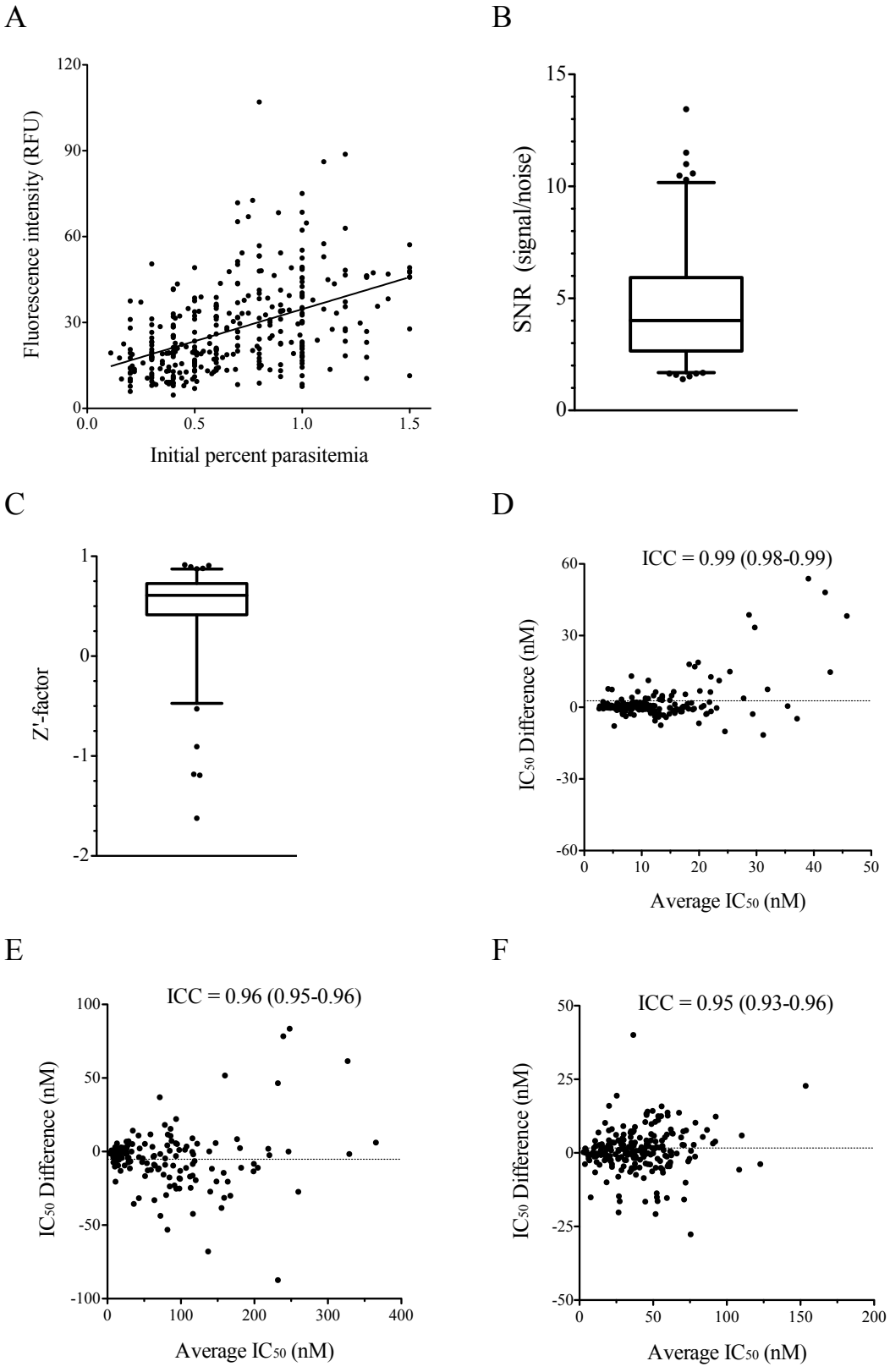
Reliability of the DAPI assay was measured by evaluating agreement between technical replicates in the untransformed scale using the intraclass correlation coefficient (ICC, calculated using *irr* in R-2.11.1 for Windows) for agreement. We analyzed only sigmoidal curves to avoid biases due to lack of fitness of different IC_{50} curve-fitting models. Agreement between the two technical replicates for each subject was excellent (Figure 4.1D-F). For all drugs, mean differences in IC_{50} between replicates were approximately zero, as expected. Except for a few outliers, differences between replicates were small compared with the IC_{50} range of each drug.

Data Analysis

Data analysis was performed in GraphPad Prism (v5.0d, San Diego, CA) and R-2.11.1 for Windows. To monitor population drug sensitivity, IC_{50} variations over time were measured

Figure 4.1. Validation of the DAPI *ex vivo* drug assay. A) Fluorescence intensity versus initial parasitemia of maximum growth (no-drug) wells for parasites tested in the DAPI *ex vivo* assay. $\rho = 0.51$, linear slope $P < .0001$. B) Signal-to-noise ratio (SNR), and C) Z'-factor for parasites tested in the DAPI assay, shown as box plots with whiskers extending to the 2.5^{th} and 97.5^{th} percentiles. D-F) Bland-Altman plots of differences between IC_{50}s of each technical replicate vs. average IC_{50} for D) amodiaquine, E) chloroqine, and F) mefloquine. Horizontal lines indicate the mean difference in IC_{50} between replicates. Intraclass correlation coefficients (ICC, with corresponding 95% confidence intervals in parentheses) are displayed on each graph. To aid data visualization, two outlier points are not shown in panels D and F and three outlier points are not shown in panel E.

Figure 4.1 (Continued)



through linear regression with $\log_{10}$-transformed IC_{50}s. Primary analysis focused on non-linear trends (using indicator variables for years) but results were confirmed by assessing linear trends. We used multiple regression models to measure whether IC_{50}s changed significantly over time after accounting for the effect of potential confounders. We also evaluated associations between IC_{50}s and possible predictors such as parasitemia (natural logarithm-transformed). Percentage changes in IC_{50} were calculated by exponentiating regression coefficients, subtracting one, and multiplying the result by 100. The equivalent calculation, using clonality of infection as an example, would be 100 x [IC_{50} (polyclonal infection) - IC_{50} (monoclonal infection)] / IC_{50} (monoclonal infection). For age, ln(parasitemia) and hematocrit, we evaluated the linearity of the association by a generalized additive model. For each individual covariate the results of all models were comparable, therefore we report the results including all covariates.

We tested whether the prevalence of any drug resistance marker changed over time by Pearson X^2 (or Fisher's exact test, as appropriate). We assessed whether there was an association between IC_{50} and the occurrence of drug resistance-associated mutations in *pfcrt* and *pfmdr1* through the Wilcoxon rank-sum test. Finally, we assessed whether IC_{50}s measured *ex vivo* were comparable to *in vitro* IC_{50}s from culture-adapted parasites by calculating the intraclass correlation coefficient (ICC) for consistency, and by linear regression.

Results

Distribution of Parasite Drug Responses Across Years

The purpose of malaria drug resistance monitoring is to evaluate parasite response to antimalarial drugs and/or the prevalence of resistance mutations, and to examine changes in a population over time. We tested *P. falciparum* drug response in patient samples collected over

three years from a clinic in Thiés, Senegal using a DAPI-based *ex vivo* assay [11]. Parasite drug responses measured each year are shown in Figure 4.2 and are summarized in Table 4.2. All parasites tested except one had amodiaquine IC_{50}s less than 70nM, suggesting that sensitivity to this drug is maintained in Senegal despite the fact that amodiaquine has been widely administered since 2006 [19]. The proportion of parasites resistant to chloroquine (IC_{50} greater than 100nM) was 26.5% in 2008, 11.5% in 2009 and 22.7% in 2010, indicating that chloroquine-resistant parasites continue to circulate in this area. Mefloquine IC_{50}s varied within the same range as chloroquine, from the single to hundreds of nanomolar.

Parasite drug response across years was significantly different only for mefloquine (Figure 4.2; *P* of categorical year in linear regression of $\log_{10}IC_{50} = 0.01$), however mefloquine summary statistics were similar across years (Table 4.2). Of the three drugs tested, none had a consistent trend of increasing or decreasing IC_{50} over time, suggesting that parasite responses stayed constant throughout the study period. There were notable changes from one year to the next, however: median IC_{50}s for both amodiaquine and chloroquine decreased from 2008 to 2009, but increased again in 2010. Conversely, the median IC_{50} for mefloquine was slightly higher in 2009 compared to both 2008 and 2010.

Prevalence of Resistance-Associated Mutations Over Time

While we did not observe large temporal changes in parasite drug response as measured by the DAPI *ex vivo* assay, the prevalence of resistance-associated mutations in both *pfcrt* and *pfmdr1* did change between 2008 and 2010 (Figure 4.3). The mutant *pfcrt* haplotype at amino acids 72-76 as well as A220S were found in more than 50% of patient samples in 2008, and while not significant, both mutations decreased between 2008 and 2010. Conversely, the *pfcrt* N326S and I356T mutations were much less prevalent, and both mutations increased slightly

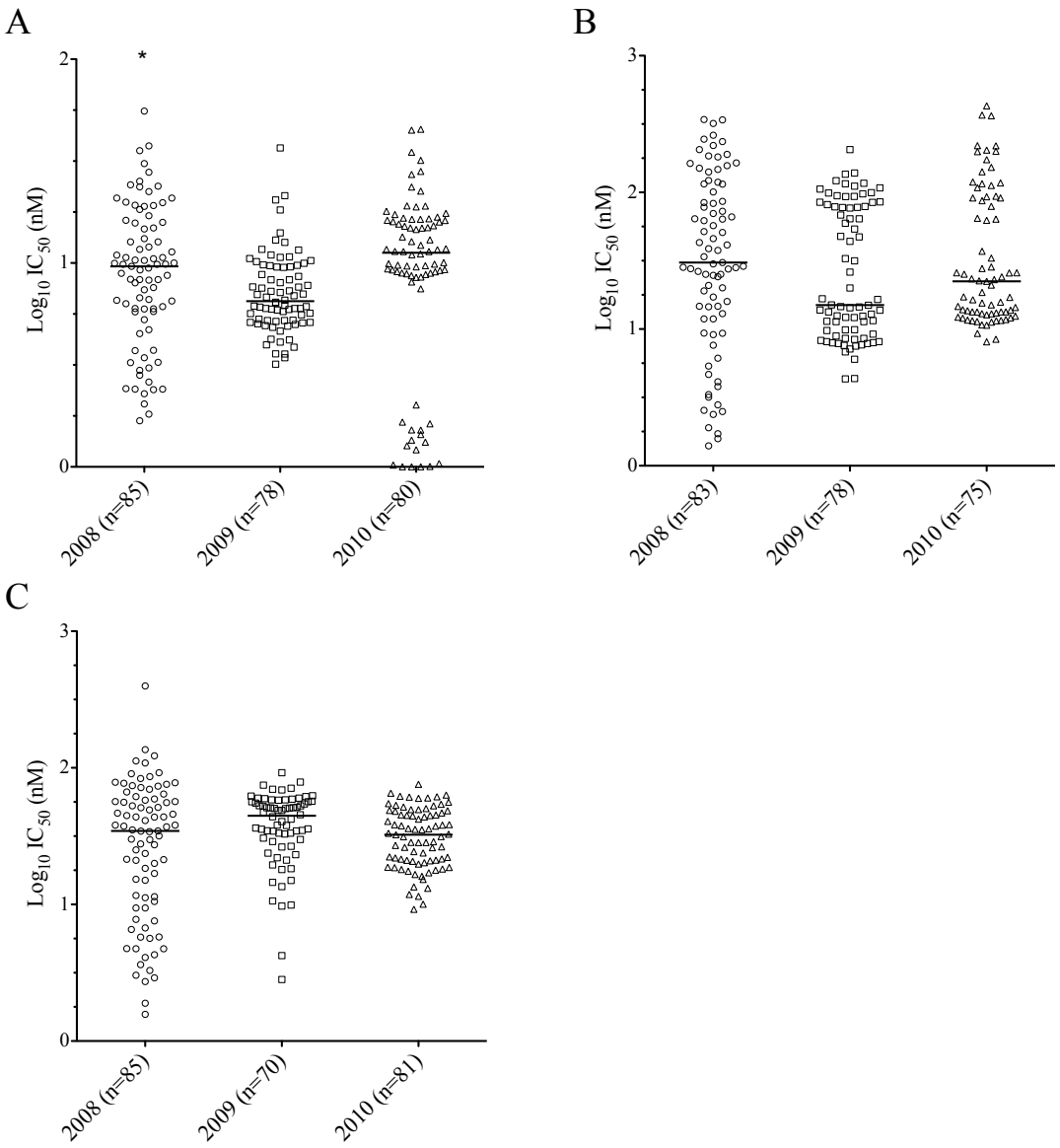


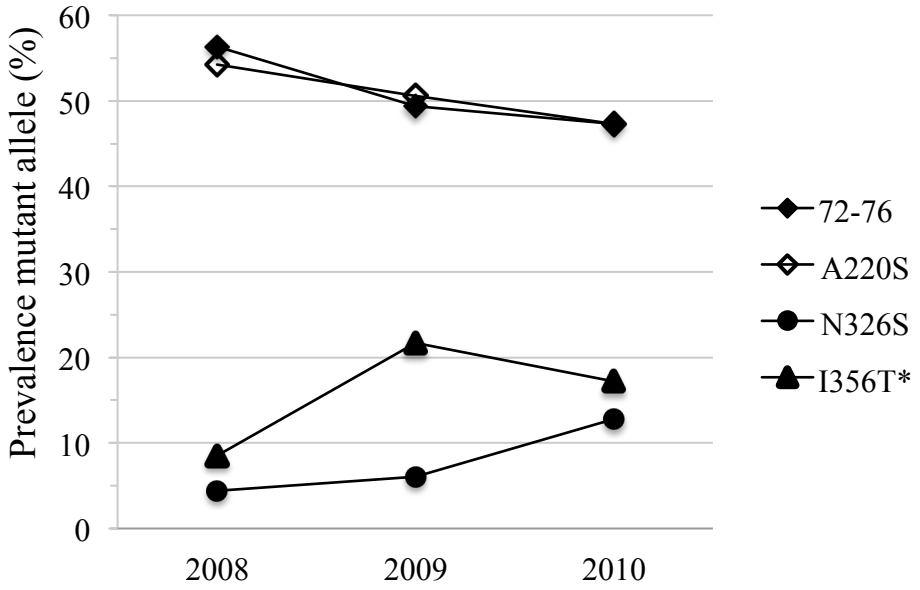
Figure 4.2. IC_{50}s in Senegalese parasites tested with the DAPI *ex vivo* drug assay in 2008, 2009 and 2010. Scatter plots show log_{10}-transformed IC_{50}s for A) amodiaquine, B) chloroquine, and C) mefloquine. Horizontal lines indicate median IC_{50}s for each year. The asterisk in panel A indicates an IC_{50} off-scale (amodiaquine IC_{50} = 1140nM). *P*-values of categorical year in linear regression of $log_{10}IC_{50}$ were 0.14 for amodiaquine, 0.43 for chloroquine, and 0.01 for mefloquine. *P*-values for linear trend were 0.52 for amodiaquine, 0.92 for chloroquine, and 0.12 for mefloquine.

Table 4.2. Parasite Drug Responses Measured in the DAPI *Ex Vivo* Assay

	2008	2009	2010
Amodiaquine			
Number	85	78	80
Median IC_{50}	9.6	6.5	11.2
Interquartile Range	5.8, 15.8	11.9, 114.9	10.0, 58.6
Percent resistant	4.7%	5.1%	5.0%
Chloroquine			
Number	83	78	75
Median IC_{50}	30.7	15	22.4
Interquartile Range	5.2, 9.6	9.1, 79.5	28.3, 56.6
Percent resistant	26.5%	11.5%	22.7%
Mefloquine			
Number	85	70	81
Median IC_{50}	34.5	44.6	32.6
Interquartile Range	8.5, 16.3	13.0, 92.0	20.8, 48.1
Percent resistant	2.4%	1.4%	1.2%

IC_{50} values are in nM. Resistance cut-offs were 100nM for chloroquine, and mean plus 2 standard deviations within each year for amodiaquine and mefloquine.

A



B

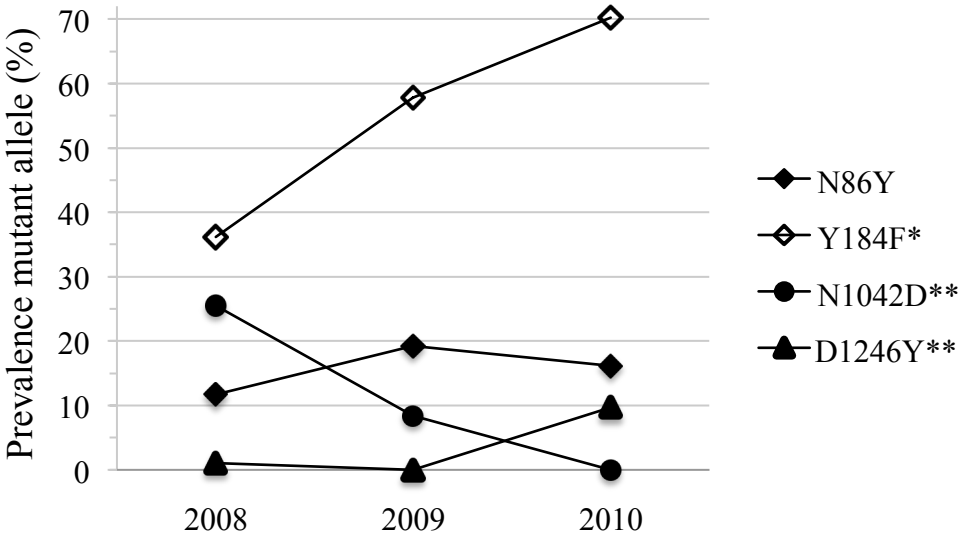


Figure 4.3. Prevalence of resistance-associated mutations in A) *pfcrt*, and B) *pfmdr1* over time, measured by high-resolution melt (HRM) technology. Prevalence was calculated by dividing the number of samples containing at least one mutant allele by the total number of samples tested each year. *$P \leq 0.05$ by Pearson X^2. **$P < .0001$ by Fisher's exact test.

between 2008 and 2010 (Pearson X^2 $P = 0.08$ and $P = 0.047$ for N326S and I356T, respectively). Prevalence of the Y184F and D1246Y mutations in *pfmdr1* also increased over time (Pearson X^2 $P < .0001$ and Fisher's exact test $P < .0001$, respectively) however the prevalence of N1042D decreased over time (Fisher's exact test $P < .0001$) and the prevalence of N86Y did not change.

Correlation Between Resistance Mutations and *Ex Vivo* Drug Response

We next looked for associations between resistance markers and *ex vivo* parasite drug responses (Table 4.3). Increased amodiaquine IC_{50} was associated with the mutant *pfcrt* haplotype at amino acids 72-76 as well as *pfcrt* A220S and *pfmdr1* D1246Y, however the differences in median IC_{50} between samples with and without each mutation were small. Chloroquine resistance was strongly associated with all the mutations we genotyped in *pfcrt*, as well *pfmdr1* N86Y and D1246Y. Increased mefloquine sensitivity was significantly associated only with *pfmdr1* N86Y and D1246Y mutations. We did not detect significant associations between responses to any of the three drugs and *pfmdr1* Y184F or N1042D.

Association Between Clinical Parameters and *Ex Vivo* Drug Response

We wanted to know whether *ex vivo* drug responses were influenced by clinical variables, including clonality of infection, hematocrit, age and parasitemia. Parasite IC_{50}s for each drug were not associated with clonality, measured as either monoclonal or polyclonal by the molecular barcode [14] (Table 4.4). IC_{50}s were similarly not associated with patient hematocrit or age. In agreement with previous findings [20], the only statistically significant predictor of IC_{50} was parasitemia, which was significant in both crude and adjusted analyses and in all models, with both the percent change in IC_{50} and *P*-values similar in all models.

Correlation Between Parasite IC_{50}s *Ex Vivo* and *In Vitro*

IC_{50}s calculated *ex vivo* may be confounded by host factors and assay variability.

Table 4.3. IC_{50} Distributions of Mutant and Wild Type Alleles for Genotypes Associated with Antimalarial Drug Resistance

Allele N(wt)/N(mut)	Median IC_{50} (Interquartile Range)								
	Amodiaquine			Chloroquine			Mefloquine		
	wild-type	mutant	*P*	wild-type	mutant	*P*	wild-type	mutant	*P*
pfcrt 72-76[a] 132/138	9 (5,14)	11 (7,18)	0.01	13 (10,23)	94 (48,173)	< 0.0001	45 (21,60)	34 (20,51)	0.09
pfcrt A220S 133/137	9 (5,13)	11 (7,19)	0.001	13 (10,23)	94 (48,173)	< 0.0001	45 (21,60)	34 (21,51)	0.2
pfcrt N326S 247/21	9 (6,16)	10 (6,18)	0.7	25 (12,93)	109 (64,163)	0.01	38 (21,57)	36 (20,50)	0.5
pfcrt I356T 228/42	10 (6,16)	10 (6,16)	0.6	24 (12,79)	97 (44,119)	< 0.0001	38 (20,56)	36 (22,58)	0.8
pfmdr1 N86Y 224/41	10 (6,15)	10 (6,20)	0.4	26 (12,85)	106 (13,202)	0.02	44 (26,59)	17 (9,22)	< 0.0001
pfmdr1 Y184F 123/148	10 (6,18)	10 (6,15)	0.6	28 (12,115)	26 (13,93)	0.4	44 (21,62)	35 (21,51)	0.08
pfmdr1 N1042D 239/31	10 (6,16)	8 (5,16)	0.3	26 (12,94)	47 (15,115)	0.5	37 (21,56)	38 (18,59)	0.8
pfmdr1 D1246Y 264/10	10 (6,15)	24 (19,31)	< 0.0001	26 (12,91)	220 (202,268)	< 0.0001	38 (22,57)	16 (13,19)	0.0005

IC_{50} values are in nM. N(wt) = number of samples possessing only the wild-type allele. N(mut) = number of samples possessing at least one mutant allele. Genotypes were determined by high-resolution melt (HRM) genotyping. Analysis excludes rare genotypes, (genotypes present at less than 2.5%). *P*-values were calculated using the Wilcoxon rank-sum test.

[a] *pfcrt* protein positions 72-76 were all perfectly correlated and were analyzed as a haplotype rather than individually.

Table 4.4. Association Between Clinical Parameters and IC_{50}s Measured in the DAPI *Ex Vivo* Assay

Variable	Percent (%) Change	*P*
Amodiaquine (N = 243[a])		
Clonality (monoclonal vs. polyclonal)	-4	0.46
Hematocrit (per % unit)	0.8	0.09
Parasitemia (per ln unit)	18	< 0.001
Age	-0.1	0.59
Chloroquine (N = 236[a])		
Clonality (monoclonal vs. polyclonal)	6	0.49
Hematocrit (per % unit)	0.3	0.65
Parasitemia (per ln unit)	18	0.006
Age	-0.2	0.91
Mefloquine (N = 236[a])		
Clonality (monoclonal vs. polyclonal)	-8	0.17
Hematocrit (per % unit)	-0.8	0.07
Parasitemia (per ln unit)	18	<0.001
Age	-0.1	0.80

Association is measured by linear regression of IC_{50} with clinical and demographic variables that could impact drug response. Analyses included all variables simultaneously, including temperature; all models produced comparable coefficients and *P*-values. ln = natural logarithm.

[a] Patients with non-sigmoidal drug curves were excluded from this analysis.

Therefore, we culture adapted 16 parasites derived from monoclonal infections collected in 2009 and repeated drug testing on the culture-adapted lines. Chloroquine IC_{50}s were consistent between *ex vivo* and *in vitro* measurements (Pearson $\rho = 0.94$; ICC for consistency = 0.80) (Figure 4.4). We also observed good agreement between mefloquine IC_{50}s ($\rho = 0.86$). Associations between amodiaquine IC_{50}s measured *ex vivo* and *in vitro* were weaker, likely because all parasites tested were sensitive to the drug and the range of observed IC_{50}s was narrow.

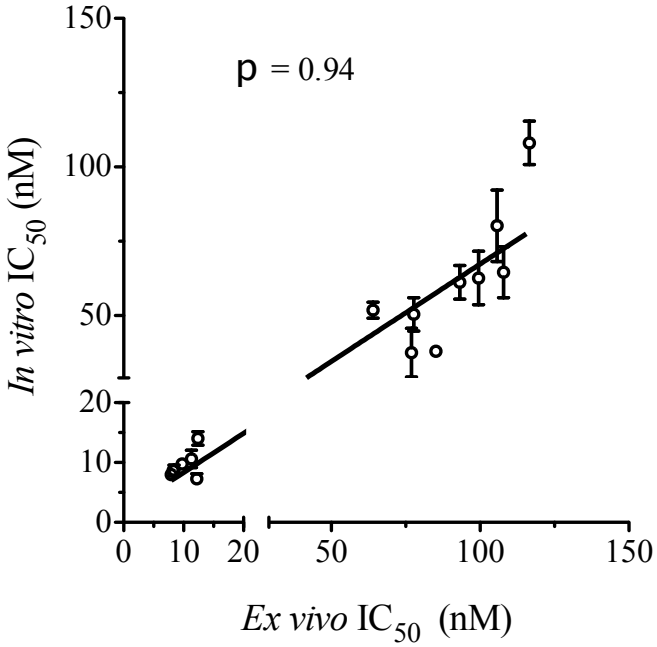


Figure 4.4. Comparison of *ex vivo* with *in vitro* IC_{50}s for chloroquine among 16 culture-adapted parasites collected in 2009. All parasites were derived from monoclonal infections. *Ex vivo* IC_{50}s were measured using the DAPI *ex vivo* assay and *in vitro* IC_{50}s were measured using a high-throughput SYBR Green I assay. Mean *in vitro* IC_{50}s are plotted with error bars showing the standard error of at least two biological replicates. ρ denotes the Pearson correlation coefficient.

Discussion

Ex vivo assays are an important tool for malaria drug resistance monitoring in direct patient samples. These assays complement *in vivo* studies by allowing researchers to test parasite response to different compounds individually and in the absence of patient factors that might introduce noise or confound results. Importantly, *ex vivo* monitoring of malaria parasite drug response can provide an early warning of decreased parasite sensitivity before parasites become highly resistant and cause infected patients to fail drug treatment.

The IC_{50}s we observed in Thiés, Senegal between 2008 and 2010 were comparable to other *ex vivo* studies of *P. falciparum* drug response [10] [21] [22]. Drug use in Senegal changed from chloroquine monotherapy to sulfadoxine, pyrimethamine and amodiaquine combination therapy in 2003, and again to ACTs in 2006 [19]. While the number of chloroquine-resistant parasites we tested decreased between 2008 and 2009, this was followed by a near doubling between 2009 and 2010, suggesting that chloroquine sensitivity may not be returning to Senegal as rapidly as it has to other areas [23] [24] [25]. Data from future years of *ex vivo* monitoring will be needed to observe how parasite drug responses continue to change in this population.

The prevalence of several resistance-associated mutations in *pfcrt* and *pfmdr1* varied between 2008 and 2011, with some mutations increasing and others decreasing in prevalence. We found a higher prevalence of the mutant *pfcrt* 72-76 haplotype than recent reports either of the mutant haplotype elsewhere in Senegal [26], or the K76T mutation in Dakar [27], but similar to these reports there appears to be a decreasing trend in mutant haplotype prevalence over time. Nonetheless, the prevalence of mutations typed within the same gene was not consistent: the 72-76 haplotype and A220S in *pfcrt* track together, but the other mutations we typed in *pfcrt* and all the mutations typed in *pfmdr1* show different prevalence from one another and inconsistent

changes over time. This suggests that selection and/or recombination may be affecting these mutations differently, and that multiple selective forces, such as use of other antimalarial drugs or other control measures, may be acting on this population.

When we looked for associations between known drug resistance mutations and parasite drug response measured *ex vivo*, we found strong correlations in almost all cases. All the genotyped *pfcrt* mutations as well as *pfmdr1* N86Y and D1246Y were associated with chloroquine resistance as measured in the DAPI *ex vivo* assay. These mutations have been previously associated with chloroquine resistance and clinical treatment outcomes [28] [29]. Conversely, we found no association between chloroquine resistance and *pfmdr1* Y184F nor N1042D, in disagreement with previous findings [30]. The Y184F mutation was not significantly associated with any of the drug we tested, but it nearly doubled in prevalence, from 36% to 70%, in the three years that we measured. Perhaps other drugs, such as quinine or artemisinin derivatives, or other selective pressures are driving this mutation to high prevalence in this population.

Ex vivo drug assays have the potential for high variability, however we found that the DAPI assay generated reliable and reproducible estimates of parasite IC_{50} in clinical samples. The dynamic range of the DAPI *ex vivo* assay was greater than that observed in previously published antimalarial drug assays using whole blood or patient samples [31] [32]. We observed strong correlation between assay technical replicates, expected associations between *ex vivo* IC_{50}s and genotyped resistance mutations, and strong correlations between *ex vivo* and *in vitro* drug responses in parasites that were culture adapted. It seems that even though *ex vivo* responses are only measured once, the DAPI assay still provides useful data.

Drug resistance monitoring using the DAPI *ex vivo* drug assay is a powerful tool for malaria control. Directly testing parasite drug responses *ex vivo* complements the genotyping of resistance mutations, and allows researchers to measure parasite response to drugs that have no known resistance markers. The DAPI *ex vivo* assay could also provide phenotype data for genome-wide analyses aimed at identifying novel drug resistance markers [33] [34] [35] [36]. As ACTs have replaced monotherapy in many areas, the ability to detect both artemisinin and partner drug resistance as they emerge will become increasingly important in order to inform treatment decisions and preserve the efficacy of these drugs.

Acknowledgments

We gratefully acknowledge members of the sample collection team at the SLAP clinic in Thiés, including Younous Diedhiou, Lamine Ndiaye, Amadou Mactar Mbaye, and Moussa Dieng Sarr. We also thank Meaghan Galligan, Justin Becker and Vishal Patel for technical assistance and thoughtful discussions.

References

1. Alonso, P.L.*, et al.* (2011) A research agenda to underpin malaria eradication. *PLoS medicine* 8, e1000406

2. Noedl, H.*, et al.* (2008) Evidence of artemisinin-resistant malaria in western Cambodia. *NEJM* 359, 2619-2620

3. Dondorp, A.M.*, et al.* (2009) Artemisinin resistance in Plasmodium falciparum malaria. *NEJM* 361, 455-467

4. Wilson, P.E.*, et al.* (2005) Real-time PCR methods for monitoring antimalarial drug resistance. *Trends in parasitology* 21, 278-283

5. Sibley, C.H. and Ringwald, P. (2006) A database of antimalarial drug resistance. *Malaria journal* 5, 48

6. Rieckmann, K.H.*, et al.* (1978) DRUG SENSITIVITY OF PLASMODIUM FALCIPARUM: An In-vitro Microtechnique. *Lancet* 311, 22-23

7. Desjardins, R.E.*, et al.* (1979) Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrobial agents and chemotherapy* 16, 710-718

8. Baniecki, M.L.*, et al.* (2007) High-throughput Plasmodium falciparum growth assay for malaria drug discovery. *Antimicrobial agents and chemotherapy* 51, 716-723

9. Bennett, T.N.*, et al.* (2004) Novel, rapid, and inexpensive cell-based quantification of antimalarial drug efficacy. *Antimicrobial agents and chemotherapy* 48, 1807-1810

10. Rason, M.A.*, et al.* (2008) Performance and reliability of the SYBR Green I based assay for the routine monitoring of susceptibility of Plasmodium falciparum clinical isolates. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 102, 346-351

11. Ndiaye, D.*, et al.* (2010) A non-radioactive DAPI-based high-throughput in vitro assay to assess Plasmodium falciparum responsiveness to antimalarials--increased sensitivity of P. falciparum to chloroquine in Senegal. *AJTMH* 82, 228-230

12. Druilhe, P.*, et al.* (2001) A colorimetric in vitro drug sensitivity assay for Plasmodium falciparum based on a highly sensitive double-site lactate dehydrogenase antigen-capture enzyme-linked immunosorbent assay. *The American journal of tropical medicine and hygiene* 64, 233-241

13. Noedl, H.*, et al.* (2004) A histidine-rich protein 2-based malaria drug sensitivity assay for field use. *The American journal of tropical medicine and hygiene* 71, 711-714

14. Daniels, R.*, et al.* (2008) A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. *Malaria journal* 7, 223

15. Daniels, R.*, et al.* (2012) Rapid, field-deployable method for genotyping and discovery of drug-resistance single nucleotide polymorphisms in Plasmodium falciparum. *Antimicrobial agents and chemotherapy*

16. Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture. *Science* 193, 673-675

17. Plouffe, D.*, et al.* (2008) In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proceedings of the National Academy of Sciences of the United States of America* 105, 9059-9064

18. Zhang, J.H.*, et al.* (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of biomolecular screening* 4, 67-73

19. WHO (2010) Focus on Senegal. In *Roll Back Malaria Progress and Impact Series*, pp. 1-56, WHO

20. Duraisingh, M.T.*, et al.* (1999) Inoculum effect leads to overestimation of in vitro resistance for artemisinin derivatives and standard antimalarials: a Gambian field study. *Parasitology* 119 (Pt 5), 435-440

21. Fall, B.*, et al.* (2011) Ex vivo susceptibility of Plasmodium falciparum isolates from Dakar, Senegal, to seven standard anti-malarial drugs. *Malaria journal* 10, 310

22. Pascual, A.*, et al.* (2012) Ex vivo activity of the ACT new components pyronaridine and piperaquine in comparison with conventional ACT drugs against isolates of Plasmodium falciparum. *Malaria journal* 11, 45

23. Mita, T.*, et al.* (2003) Recovery of chloroquine sensitivity and low prevalence of the Plasmodium falciparum chloroquine resistance transporter gene mutation K76T following the discontinuance of chloroquine use in Malawi. *The American journal of tropical medicine and hygiene* 68, 413-415

24. Kublin, J.G.*, et al.* (2003) Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. *The Journal of infectious diseases* 187, 1870-1875

25. Laufer, M.K.*, et al.* (2010) Return of chloroquine-susceptible falciparum malaria in Malawi was a reexpansion of diverse susceptible parasites. *The Journal of infectious diseases* 202, 801-808

26. Ndiaye, M.*, et al.* (2012) Assessment of the Molecular Marker of Plasmodium falciparum Chloroquine Resistance (Pfcrt) in Senegal after Several Years of Chloroquine Withdrawal. *The American journal of tropical medicine and hygiene*

27. Wurtz, N.*, et al.* (2012) Prevalence of molecular markers of Plasmodium falciparum drug resistance in Dakar, Senegal. *Malaria journal* 11, 197

28. Djimde, A.*, et al.* (2001) A molecular marker for chloroquine-resistant falciparum malaria. *N Engl J Med* 344, 257-263

29. Picot, S.*, et al.* (2009) A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria. *Malaria journal* 8, 89

30. Foote, S.J.*, et al.* (1990) Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in Plasmodium falciparum. *Nature* 345, 255-258

31. Bacon, D.J.*, et al.* (2007) Comparison of a SYBR green I-based assay with a histidine-rich protein II enzyme-linked immunosorbent assay for in vitro antimalarial drug efficacy testing and application to clinical isolates. *Antimicrobial agents and chemotherapy* 51, 1172-1178

32. Vossen, M.G.*, et al.* (2010) The SYBR Green I malaria drug sensitivity assay: performance in low parasitemia samples. *The American journal of tropical medicine and hygiene* 82, 398-401

33. Mu, J.*, et al.* (2010) Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nature genetics* 42, 268-271

34. Van Tyne, D.*, et al.* (2011) Identification and Functional Validation of the Novel Antimalarial Resistance Locus PF10_0355 in Plasmodium falciparum. *PLoS Gen* 7, e1001383

35. Cheeseman, I.H.*, et al.* (2012) A major genome region underlying artemisinin resistance in malaria. *Science* 336, 79-82

36. Park, D.J.*, et al.* (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*

DISCUSSION

Shall we dance?

Human and malaria evolution in an era of eradication

Note: This essay was composed by Daria Van Tyne and was submitted for the Bowdoin Prize, Harvard University

Summary

Humans are intertwined in a complex evolutionary dance with *Plasmodium falciparum*—the parasite responsible for the fatal disease we know as malaria. Today, our sights are set on malaria eradication, yet achieving this goal requires that we first appreciate the resilience of our relationship with this important parasite.

Falcutate magis quam violentia.[1]

- Hippocrates

Dynamics of a disease

The child sits limply on her mother's lap, her lips parted, eyes half closed. Outside the clinic, the hot African sun takes no prisoners—it beats down unrelentingly on the dirt roads of the village, the crops already parched though the rains have barely ceased, the tin roof and clay walls of the clinic. Mother and child left home before dawn in order to reach the clinic before the sun rose high and the child's fever came; now she is as hot as the sun, listless and uninterested in the bustle of patients moving around the busy clinic. She doesn't flinch when her finger is pricked. A single drop of blood reveals that she has malaria, as both her mother and the doctor suspected. Inside her bloodstream, the parasites are multiplying so quickly that they can destroy one-quarter or more of her red blood cells in just two days. Without treatment, the prognosis is quite grim. But with the proper medication, the parasites will disappear from her bloodstream in a matter of hours, and she will come out of her daze to resume her normal, day-to-day life. That is, until the next mosquito bite.

Imagine this same scene played out every eight seconds, which is about as often as someone becomes sick with malaria somewhere in the world.[2] Sometimes the child in this story gets better; sometimes, for any number of reasons, they do not. The child might fall into a coma if the parasites adhere to the small vessels in their brain and block the blood from passing

[1] "By skill, rather than by violence."

[2] "World Health Organization fact sheet on malaria." http://www.who.int/mediacentre/ factsheets/fs094/en (accessed November 5, 2010).

through. They might succumb to anemia and septic shock caused by parasites that literally eat up red blood cells in droves. Sometimes the patient is not a child but rather a pregnant woman in Mozambique, who comes perilously close to losing her baby's life, as well as her own, when malaria parasites invade the placenta of a first-time mother. Sometimes the patient is a rural forest worker in Thailand or Brazil, who makes his living cutting down trees in the jungle until he is stricken with a multi-drug-resistant form of malaria that responds to treatment at first but comes roaring back again weeks or months later.

These scenarios share the common thread of a terrible disease, which is both preventable and treatable, yet still remains a formidable obstacle to global health. In this essay, I aim to build an understanding of the relationship that has formed between humans and the malaria parasite. I view this relationship as something akin to a dance, with both partners influencing and responding to the actions of the other. A global campaign aimed at eradicating malaria is now underway, however this campaign will only be successful if we can first understand the dance – its tempo and rhythms, its movements and steps – that has made malaria such a remarkably successful infectious disease. This understanding is essential because it will form the step that has been missing in the past, which may allow us to finally bring the dance to a close.

Setting the scene

Our engagement with *Plasmodium falciparum* stretches back so far that it is difficult to know precisely when we, as a species, first became infected. We would undoubtedly prefer to be rid of malaria; *P. falciparum* is a parasite that causes nearly a million deaths every year.[3] But this unique organism has responded to our attempts to eliminate it by adapting to each advance,

[3] "World Health Organization fact sheet on malaria." http://www.who.int/mediacentre/ factsheets/fs094/en Accessed November 5, 2010.

evolving ways to escape the human immune system and resistance to anti-malarial drugs. Like any good dance partner, we lead and it follows. Conversely, malaria has had a role in shaping our own evolutionary history, by preserving various red blood cell disorders that together have formed unmistakable malaria "fingerprints" on the human genome. As the tempo quickens and we now seek to eradicate malaria parasites, it is essential that we first understand the nature of our dance with these cunning and crafty creatures.

The vast majority of malaria deaths today occur in children under five years old and pregnant women living in Sub-Saharan Africa. However, nearly half of the world's population – 3.3 billion people – is at risk of being infected with *Plasmodium* parasites.[4] In Africa, malaria accounts for nearly 20% of all childhood deaths and is the largest single killer of children on the continent. But malaria is a tremendous problem in other parts of the world as well: Southeast Asia is a hotbed for drug-resistant parasite strains, which first surfaced in the jungles of Thailand and Cambodia and quickly spread worldwide. In Latin America, as many as 3 million people are chronically infected with malaria;[5] this high morbidity takes its toll not only on those infected but also on their families and local communities, through lost days of work and diminished productivity.

Clearly, malaria is an enormous global health problem. A renewed effort is currently underway to control malaria, and ultimately to eradicate it from the face of the earth. National governments, private donors, and researchers worldwide have issued calls to work together and "close the book" on this fatal disease. Of course, malaria eradication has been tried in the past. In the 1950s and 60s, the World Health Organization mobilized a global effort aimed at malaria

[4] "Malaria Key Facts." http://www.rollbackmalaria.org/keyfacts.html (accessed November 5, 2010).

[5] Mendis et al. (2001). The neglected burden of *Plasmodium vivax* malaria. *American Journal of Tropical Medicine and Hygiene* **64**: 97-106.

eradication; its director Marcolino Candau confidently declared in 1955: "We have the tools; we know what needs to be done; it is [now] simply a matter of going out and doing it." Unfortunately, the only thing that this global campaign eradicated was *interest* in malaria, and a lack of sustained effort led to the eventual resurgence of the disease. The failure to eliminate malaria during the 20th century was likely in part due to a lack of understanding of the step/counterstep dynamic of our relationship with this parasite, whereby the forces that were applied toward disease eradication resulted in genetic adaptations that allowed malaria parasites to overcome those forces and thrive once again. The current malaria eradication effort can be successful only if scientists and citizens alike first understand how our uneasy dance with the malaria parasite came to be. We must understand how we have shaped malaria's evolutionary history and, conversely, how it has shaped our own evolution.

Malaria 101: the basic steps

The disease "malaria" refers to infection with any one of five distinct species that scientists and doctors refer to as *Plasmodium*. *Plasmodium falciparum,* the major focus of this essay, is the most well known parasite of the group because it is the most widespread and most deadly form of malaria. But malaria can also be caused by any of four other species: *Plasmodium vivax, Plasmodium malariae, Plasmodium ovale*, or *Plasmodium knowlesi*, the last of which has only recently been discovered to infect humans.[6] Malaria caused by one of these other parasites is generally milder than a *P. falciparum* infection. Nevertheless, chronic infection with a non-*falciparum* species, like *P. vivax,* is a major contributor to the morbidity caused by *Plasmodium* parasites.

[6] Cox-Singh et al. (2008). *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis.* **46**(2):165-71.

P. falciparum is a single-celled protozoan: it has a nucleus and is more biologically complex than a bacterium, but it remains fundamentally different from other unicellular organisms such as yeast, and equally different from the cells that make up plants and animals. The *P. falciparum* genome—all the DNA within the organism—contains 23 million molecules of DNA, which encode instructions for how to produce and control over 5,000 genes.[7] This large number of genes makes malaria among the more complex of human infectious agents. By contrast, viruses typically have only 10 to 15 genes, while bacteria have 500-2,000 genes.

Plasmodium parasites were first described in the late 19th century by Alphonse Laveran, a French army doctor in Algeria, who observed strange-looking organisms inside the red blood cells of a soldier suffering from anemia and fever.[8] However, the first description of the symptoms of malaria is much older; it is found in Chinese medical writings that date from about 2700 B.C.E.[9] In ancient Greece, Hippocrates observed patients bearing the salient symptoms of malaria. Noting the regular fevers and powerful chills that caused this "bad air," Hippocrates correctly presumed that the disease was in some way associated with the marshes that existed around Athens, which harbored large mosquito populations in the summer and fall.[10] The malaria parasite was even found in King Tut's remains from 1324 B.C.E., and researchers recently declared that the young pharaoh likely died in part because of a severe bout of malaria caused by *P. falciparum*.[11] King Tut is in pretty good company, though: other famous figures that are likely

---

[7] Gardner et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**(6906): 498-511.

[8] Laveran A. (1893). *Paludism: marsh fever and its organism*. New Sydenham Society Publications 146, London.

[9] Man (1950) Nei Ching, the Chinese Cannon of Medicine. *Chinese Medical Journal* **68**(1-2): 1-33.

[10] Cunha and Cunha (2008). Brief history of the clinical diagnosis of malaria: from Hippocrates to Osler. *Journal of Vector Borne Diseases* **45**: 194-9.

to have died from malaria infection include Alexander the Great, Oliver Cromwell, Genghis Khan, and Dante Alighieri.[12]

To complicate matters, the uneasy dance in which we're locked with malaria is actually something of a threesome. For without female *Anopheles* mosquitoes acting as the middleman, *P. falciparum* could not be passed from one person to another. Inside the human bloodstream, malaria parasites replicate asexually, making each daughter parasite a clone of its mother. But deep within the belly of a mosquito, malaria takes up a sexual life cycle, whereby different parasites are able to swap genetic material and generate unique offspring.

The malaria parasites that we find throughout the world today are highly diverse, meaning that parasites can belong to the same species but have very different DNA. There are tens of thousands of elementary genetic differences—scientists call them single nucleotide polymorphisms ("SNPs") or mutations—which distinguish one parasite from another, much in the same way that people possess different traits such as hair or eye color, which are genetically encoded by different versions of the same genes. This happens in malaria too, but the traits are less aesthetic and have more to do with things like how well the parasite infects a red blood cell or whether it will be recognized by the human immune system and removed before it makes a person feel sick.

The elementary genetic differences that make each parasite (and each human) unique are the basis for the process of biological evolution. From one generation to the next, nature randomly introduces mutations into the genome. Usually these mutations are not beneficial, but in rare instances, a random mutation gives the individuals bearing it a comparative advantage

---

[11] Hawass et al. (2010). Ancestry and pathology in King Tutankhamun's family. *Journal of the American Medical Association* **303**(7): 638-47.

[12] "Malaria Site: History of Malaria: Famous Victims." http://www.malariasite.com/malaria/ history_victims.htm (accessed November 11, 2010).

over their peers. Advantageous mutations generally confer a greater chance of survival and reproductive success in a particular environment, and both of these ensure that a beneficial mutation will be preserved and passed on to the next generation. As a result, advantageous mutations can propagate rapidly through a population in just a few generations.

To increase the chances of hitting upon one of those rare but beneficial mutations, humans rely on meiosis. This is a unique process of cell division that mixes and matches maternal and paternal DNA and results in gametes that, when combined, produce a progeny that is much more diverse than a mere carbon copy of either parent would be. *Plasmodium* parasites, by contrast, achieve a similar result through a slightly different process. Asexually-replicating parasites in the human bloodstream are sensitive to environmental triggers, such as anemia or low blood sugar, which spur them to differentiate into sexual forms of the parasite. Biologists refer to these forms as "gametocytes." When a mosquito ingests gametocytes, these develop further into male and female gametes, and the gametes subsequently combine to produce a unique zygote.[13]

The sexual stage of the malaria life cycle allows for the shuffling of mutations that will ultimately prove beneficial or deleterious to parasite progeny, depending on the nature of the mutations and the environment that the progeny encounter. For example, a mutation that confers resistance to a commonly used anti-malarial drug would give the progeny bearing that mutation the ability to survive inside the bloodstream of people who have taken the drug. However, the same drug would kill their "brother" and "sister" parasites that lack the mutation, effectively eliminating their genes from the gene pool. This "natural selection" of organisms that are able to survive in a particular environment is familiar to us as evolution. When viewed from the

[13] Talman et al. (2004). Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malaria Journal* **3**: 24.

standpoint of malaria's relationship to its human hosts, it could also be conceived as an uneasy dance occurring between two hostile partners.

The beat goes on: malaria evolution

Let us now take a closer look at the elaborate dance linking us with *P. falciparum*, and in particular, how malaria can "follow our lead" and adapt to the unwelcoming environments it encounters inside the human body.

Imagine that you are a malaria parasite. By necessity, you must live inside a person, which will involve some pretty intricate footwork. The human immune system is a complex and sophisticated machine designed to identify and destroy invaders, just like you. Additionally, the immune system is designed to "remember" each and every invader, so that the second time a particular pathogen tries to establish an infection, the body's defenses are already primed against it. The most successful pathogens have figured out ways to bypass the immune system and thrive inside their human hosts—despite the body's best attempts to clear the infection. One particularly successful strategy for immune evasion is to constantly "change outfits," so that the recognition machinery of the human immune system always thinks that you are a brand-new organism that it has never seen before.

Malaria parasites do just this. The immune system normally identifies pathogens by the proteins that they express at surface level. By periodically switching surface proteins (namely, different versions of a protein called *var*), *P. falciparum* parasites are able to hide from the normal actions of the human immune system. To understand how this is achieved, note first that the malaria genome contains all of the information necessary to keep the parasite alive, including all of the "recipes" for each protein that it will need during its lifecycle. The *P. falciparum* genome includes at least 60 different copies of the gene that contains the *var* protein recipe. The

vars are also highly polymorphic, meaning that the protein sequence of the same *var* varies dramatically between different parasites. The ability to "switch" rapidly, displaying distinct versions of the *var* protein in succession, is what gives the parasite a leg up on our immune system. It takes five to seven days for the human body to mount an effective immune response to a pathogen that it has encountered. Since the *P. falciparum* life cycle in the human bloodstream is only 48 hours, the parasite has more than enough time to "change outfits" and switch to a different *var* protein that the immune system no longer recognizes.

In addition to the *var* genes, *P. falciparum* contains several other groups of polymorphic genes that are all dedicated to the task of keeping the human immune system at bay. This kind of constant variation is not unique to malaria parasites; the switching of surface proteins (also known as antigens) in the influenza virus is the reason why people need to get a flu shot each year, as it would be impossible to make a vaccine that would protect against every possible flu strain. In the case of flu, a new vaccine is made each year that protects against the most common circulating virus strains. It's possible that this approach could eventually be applied to malaria, though at the moment there is no easy way to identify which parasite strains are circulating in a particular area. Furthermore, because *P. falciparum* has 50-fold more genes than the influenza virus, it is likely that there are simply too many distinct parasite strains circulating in malaria-endemic areas for this approach to be effective.

With its capacity for antigenic variation, the malaria parasite clings to us like a shadow, despite the human immune system's best efforts to break from its embrace. Indeed, recent analysis of the *P. falciparum* genome has revealed that the most highly diverse genes (i.e., the genes with the greatest number of mutations) in the entire *P. falciparum* genome are those that

encode variable surface proteins.[14] This suggests that the strongest force for the selection of beneficial mutations in the parasite—so-called "positive selection"—is the human immune system itself.

After the human immune system, the second strongest type of selective pressure that we have placed on the malaria parasite involves the use of anti-malarial drugs. Interestingly, two of the most effective medicines for the treatment of malaria, artemisinin and quinine, are also the two oldest known anti-malarials. Artemisinin, extracted from the wormwood *Artemisia annua*, was first used in China as early as 168 B.C.E., and forms a crucial part of the currently recommended treatment for *P. falciparum* malaria, known as artemisinin combination therapies or ACTs. Quinine was discovered during the early 17th century by Jesuit missionaries in Latin America. After learning from indigenous tribes that tea made from the bark of the *Quinquina calisaya* tree could cure their dreaded fevers, the missionaries quickly shipped the bark back to Europe for the treatment of malaria there. In the mid-19th century, British colonists in India realized that quinine-infused tonic water could be combined with gin to make a refreshing beverage that also happened to have anti-malarial properties.[15] Today, quinine is still widely used, and it remains the only recommended treatment for patients suffering from severe malaria. Artemisinin and quinine are not the only drugs that have been used to treat malaria, but unfortunately, the vast majority of anti-malarials have been rendered ineffective by the rise of drug-resistant *P. falciparum* strains.

Incidentally, the very same anti-malarial drug that gave scientists in the 1950s the confidence to declare malaria eradication possible failed because it drove the evolution of drug

---

[14] Volkman et al. (2007). A genome-wide map of diversity in *Plasmodium falciparum. Nature Genetics* **39**(1): 113-9.

[15] "The History of Gin and Tonic." http://www.bbc.co.uk/dna/h2g2/alabaster/A568677 (accessed November 16, 2010).

resistant parasites. During the First World War, Germany was unable to obtain quinine from Java, where it was harvested, and as a result, German soldiers had no way to properly treat the malaria that many of them contracted while fighting in Southern Europe.[16] This sparked the commission of a group of pharmaceutical chemists tasked with searching for a synthetic alternative to quinine, which at the time was the only available anti-malarial in Europe. The chemists came up with several alternative compounds that were all derivatives of quinine. One of these derivatives was further modified by American chemists, and this compound, named chloroquine, was widely hailed as a new, safe, and remarkably effective anti-malarial drug.[17]

Cheap and highly effective at killing malaria parasites, chloroquine was the first (and last) anti-malarial drug to be mass administered to entire populations that were at risk of malaria infection in the tropics. Just as iodine has been added to table salt for almost a century in order to protect against deficiencies in this essential element, chloroquine was also added to table salt beginning in the 1950s in an attempt to "pre-treat" populations in malaria-endemic areas.[18] Chloroquine's low cost and lack of toxic side effects made it an easy choice for prophylactic treatment—but soon, the malaria parasite mutated and chloroquine-resistant parasites rapidly developed in both Southeast Asia and South America. From there, resistance was quick to spread to Africa, and by the 1980s, chloroquine was rendered completely ineffective as a worldwide treatment for malaria.

With our newly developed ability to scan the malaria parasite's genome for evidence of recent selective pressures, we are able to peer into malaria's evolutionary history. It is now

---

[16] Butler et al. (2010). A brief history of malaria chemotherapy. *The Journal of the Royal College of Physicians of Edinburgh* **40**(2): 172-7.

[17] Surrey et al. (1946). Some 7-substituted 4-aminoquinoline derivatives. *Journal of the American Chemical Society* **68**: 113-6.

[18] Payne (1988). Did medicated salt hasten the spread of chloroquine resistance in *Plasmodium falciparum*? Parasitology Today **4**(4): 112-5.

possible to uncover the tell-tale fingerprints of recent pressures that we have imposed on the parasite population through the use of anti-malarial drugs, particularly chloroquine. The emergence of chloroquine-resistant parasites is almost entirely due to mutations in a gene that biologists call the *chloroquine resistance transporter*. This gene encodes a cellular pump that, when mutated, is able to quickly remove poisonous chloroquine from the parasite, allowing the mutant parasites to continue growing even in the presence of high concentrations of the drug.[19] An analogous mutation appeared when patients were treated with mefloquine, which selected for parasites bearing mutations in a separate transporter gene. Likewise, treating patients with pyrimethamine was met by resistance mutations in a different gene called *dihydrofolate reductase*.

The list of gene mutations goes on and on. Indeed, drug resistant parasites have developed in response to every single anti-malarial drug that has been used to treat infected patients. Every new drug (or step) that we try has resulted in a counterstep from the parasite; namely, the appearance of parasites bearing mutations that make them resistant to that drug, sometimes within as little as a few months. We now know that when a large population of pathogens (whether bacteria, viruses or parasites) is treated with a single drug for a long time, it is possible to select one "special" individual that has a critically beneficial mutation. This individual alone is able to survive in the face of the drug pressure that has been imposed. It can multiply and repopulate the environment, spreading its resistance mutation or mutations far and wide. For this reason it is critically important to monitor the efficacy of the anti-malarial drugs used in a particular area, and to discontinue their use in favor of other drugs if it appears that the parasite population has developed resistance.

---

[19] Martin et al. (2009). Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* **325**(5948): 1680-2.

Anti-malarial drug resistance among *P. falciparum* parasites was one of the major contributors to the failed malaria eradication campaign of the 1950s and 1960s. In 2004, the World Health Organization finally acknowledged that any anti-malarial drug that was given alone would likely generate drug-resistant parasites. This so-called "monotherapy" was banned, and the Organization began instead to recommend treatment through *combinations* of drugs. Usually this consists of a form of artemisinin (which while effective, is rapidly broken down in the bloodstream) combined with a long-lasting partner drug. Currently, artemisinin combination therapies (ACTs) are the most widely used and most effective treatment for malaria. But a reduction in the efficacy of ACTs has already been observed on the Thai-Cambodian border, which has historically been a hotspot for drug-resistant malaria strains.[20] It seems that it may only be a matter of time before this latest step—the move to combination therapies—results in parasites that are resistant to every possible drug combination that we can use.

Complicated choreography: malaria's imprint on human evolution

It is clear that both the human immune system and our use of anti-malarial drugs have had a dramatic impact on the malaria genome: we lead and the parasite follows, adapting and responding to our every move. But like a well-chosen dance partner, *P. falciparum* is also able to mirror our steps with its own, and in this way, malaria has had a marked effect on the evolutionary history of humans. In the same way that our immune systems and anti-malarial drug use have selected for the parasites that are most "fit" to survive in a particular environment, so too has *P. falciparum* exerted a strong selective pressure on human evolution.

[20] Noedl et al. (2008). Evidence of artemisinin resistant malaria in western Cambodia. *New England Journal of Medicine* **359**: 2619-20.

In the midst of the global malaria eradication effort of the 1950s, doctors and researchers started to notice a curious coincidence: many human red blood cell disorders are most common in precisely the same regions of the world where infection with malaria is most common—the tropics and subtropics.[21] Sickle cell anemia is the classic example of this phenomenon. Among some African tribes, as much as 40% of the population carries sickle cell mutations in either one or both copies of their *hemoglobin* gene.[22] Hemoglobin is the molecule inside red blood cells that delivers oxygen to all the cells in the body, and sickle cell mutations result in misshapen hemoglobin that is less effective at transporting oxygen to body tissues. Here a puzzle arises: if sickled red cells are less effective at delivering oxygen and are thus detrimental to humans, how can it be that evolution has not removed these deleterious mutations from the gene pool?

To understand how this can happen, imagine an African man and woman who both carry the sickle cell trait; that is, they each have one normal copy of the hemoglobin gene and one defective copy. If this couple has four children, the odds are that one of their children will inherit both copies of the sickle-cell hemoglobin gene (one from each parent) and will have a high risk of dying from sickle cell anemia, since the child's red blood cells will be unable to transport oxygen efficiently to their body tissues. Another child is likely to receive two normal copies of the hemoglobin gene, and will be perfectly healthy—that is, until they are infected with *P. falciparum* and succumb to malaria. The other two children are likely to be lucky: they would both inherit one normal copy and one sickle cell copy of the hemoglobin gene. Like their parents, they might experience mild symptoms of sickle cell anemia, such as pain in their extremities, dizziness or shortness of breath. But because they are heterozygous for the sickle trait (meaning

[21] Allison (1954). Protection afforded by the sickle-cell trait against subtertian malareal infection. *British Medical Journal* **1**(4857): 290-4.

[22] Each of us has two copies of every gene: one inherited from our mother and one from our father.

they have one normal copy and one mutated copy of the hemoglobin gene), they have a huge "fitness advantage" over their healthy sibling when it comes to malaria. If either of the two heterozygous children becomes infected with *P. falciparum*, the parasite will be unable to live within their sickled red blood cells, and they will have a much easier time surviving the infection. Because only half of their red cells are sickled, the children would still become infected with malaria, but the advantage is enough to tip the balance towards the human partner. This neatly explains why otherwise harmful mutations in the hemoglobin gene are preserved within the genomes of those at risk for malaria infection.

Earlier we considered scientists' ability to scan the malaria genome in order to learn about the various ways that we have directed malaria's evolution, both through our immune systems and more recently through the use of anti-malarial drugs. Recent analysis of the human genome has similarly revealed that many people living in malaria-endemic areas harbor identical hemoglobin mutations that cause sickle cell anemia, suggesting that they descended from a common ancestor that originally carried those mutations.[23] The hemoglobin gene is just the tip of the iceberg: numerous other disorders of red blood cells, such as dysregulation of hemoglobin protein levels and ovalocytosis (oval-shaped red blood cells) have also been preserved in populations at risk for malaria infection. Indeed, malaria has shaped our genome more dramatically than any other evolutionary force known to date. This counterbalance of two-way evolutionary pressure does not exist between us and any other parasitic disease, and it makes malaria unique in having dramatically shaped our biological history. Like an overbearing dance partner, the parasite seems to pull us closer even as we try to push away.

---

[23] Kwiatkowski (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *American Journal of Human Genetics* **77**(2): 171-92.

Exiting the floor: The "E" word

Our uneasy dance with the malaria parasite has endured for thousands of years; the constant back-and-forth has resulted in dramatic changes in both the parasite's genome and our own. As we now set our sights on malaria eradication, an appreciation for the resilience of our relationship with this parasite is critical to the understanding of how we might be able to rid humanity of this disease.

The nearly one million annual deaths caused by *P. falciparum* each year make malaria a devastating burden to global health. Yet if Bill and Melinda Gates have their way, in just a few years malaria will be extinct—eliminated from the planet at large and confined to biology research labs and textbooks. The Gates Foundation has partnered with organizations such as Roll Back Malaria, the Global Fund, and the Medicines for Malaria Venture, and together they have set the ambitious goal of completely eradicating malaria before the year 2050.[24] It is an enormous task, but one that in principle seems feasible using currently existing technologies. We have anti-malarial drugs that, when given in combination, are highly effective at eliminating parasites from the bloodstreams of infected patients. We have insecticide-treated bed nets that can prevent infectious mosquito bites. We have enough knowledge of the mosquito life cycle to realize that simple measures, like draining standing water near homes, will reduce the numbers of mosquitoes to which people are exposed. Still, the fact that eradication is possible *in principle* does not mean that the task will be straightforward.

The global malaria eradication campaign of the 1950s and 1960s was unsuccessful for three major reasons: (1) *P. falciparum* parasites became resistant to chloroquine; (2) *Anopheles* mosquitoes became resistant to insecticides, DDT chief among them; and (3) governments all over the world lost their political will and failed to keep up a sustained effort to identify new

[24] Roberts and Enserink (2007). MALARIA: Did They Really Say … Eradication? *Science* **318**(5856), 1544 - 1545.

cases and combat malaria with education, the only tool that was still effective even after drug resistance had emerged and spread. While we now know much more than we did in the 1950s about the biology of both the parasite and the mosquito, the previous failed attempt to rid humanity of the scourge of malaria should make us more than a little skeptical that this parasite will be easily brought under control.

Malaria was once common throughout the United States, and for a long time it was a particularly large problem in the American South. As Oliver Wendell Holmes observed, malaria also appeared in limited outbreaks around Lake Champlain and along numerous rivers in Massachusetts.[25] Government-led malaria eradication campaigns during the 1930s and 1940s successfully drove the parasite out of North America, primarily due to the draining of mosquito breeding habitats and through widespread use of insecticides like DDT. Malaria was successfully eliminated from temperate zones in the US and Europe because these areas had lower transmission rates compared with the tropical climates that exist in large parts of Africa and Asia. It was much easier to drive transmission below the level needed to propagate the infection when the rate of transmission was already low to begin with. These were also relatively resource-rich areas that could afford to invest in the manpower necessary to tackle the disease. Despite the hard work of numerous people over many years, *Anopheles* mosquitoes that remain capable of transmitting malaria are still found throughout the United States today. As people travel to distant places with increasing frequency, it is not difficult to imagine how one malaria-infected person would have the potential to spark a new epidemic of the disease, even in America.

Just as we now know that the most effective treatment for a person who is sick with malaria is to use a combination of anti-malarial drugs, we also understand that the most effective way to eliminate *P. falciparum* from an area is to use a combination of intervention strategies.

[25] Holmes, Oliver Wendell. *Medical Essays, 1842-1894*. Houghton, Mifflin and Company, 1891. p. 324.

This is why current efforts at malaria eradication seek to combine interventions—anti-malarial drugs to treat the sick, insecticides to kill mosquitoes, a malaria vaccine to protect from infection, and possibly even genetically-modified mosquitoes that are unable to transmit malaria parasites. Each strategy used alone would be relatively ineffective: parasites will become resistant to anti-malarial drugs, mosquitoes will become resistant to insecticides, a malaria vaccine is unlikely to fully protect against every possible parasite strain, and genetically-modified mosquitoes may not survive outside of the laboratory. But even with these limitations in mind, each control strategy targets a separate aspect of the malaria life cycle, and thus concerted efforts that employ combinations of these measures have the potential to dramatically alter malaria's impact in an endemic area. More important that any of these interventions, however, is the necessity for government leaders to continue to look for new cases of malaria for many years after control measures have been implemented. Only then can we be sure that our dance with the parasite has truly come to an end.

Consider the case of smallpox, the only human disease that has been successfully eradicated from the planet. Smallpox is a virus that only infects humans; it was eliminated in large part due to the development of an effective vaccine and a sustained effort to make sure that everyone in the entire world who could possibly be infected received the vaccine. Although *P. falciparum* is a much more complex organism, it causes a type of malaria that also only infects humans. While mosquitoes are required for malaria transmission, there is no other animal "reservoir" to which parasites can escape. In that sense, there is no option for malaria to sit out a song or two and then re-emerge to resume its dangerous dance with us. Thus it seems that combining our "tools" with enough sustained political will, the dream of the Gates Foundation is likely to one day come true.

Oliver Wendell Holmes also noted that "Nature withheld the fatal gift of malaria," as a true epidemic, from his part of the world, but "[Nature filled the world] with exhalations that breed the fever of inquiry in our blood and in our brain."[26] This "fever of inquiry" has led us to turn our eyes to the inner-workings of the malaria parasite and to explore how its evolutionary history is tied inextricably to our own. Today, as we are called once more to work towards malaria eradication, we must remember the missteps of the past and appreciate the long and dynamic history that we have with the malaria parasite. The dance has been going on for thousands of years, but an end may finally be within reach.

[26] Holmes, Oliver Wendell. *Medical Essays, 1842-1894*. Houghton, Mifflin and Company, 1891. p. 206.

APPENDIX A

Supplemental Methods, Tables and Figures for Chapter One:

Van Tyne et al. (2012). Identification and Functional Validation of the Novel Antimalarial Resistance Locus *PF10_0355* in *Plasmodium falciparum*. *PLoS Genetics* 7(4): e1001383.

Supplemental Methods

Parasites

Parasites (Table S1) were obtained from the Malaria Research and Reagent Resource Repository (MR4, malaria.mr4.org) or additional sources noted (Supplemental Methods, Table S1, Acknowledgements). The following parasite lines were obtained through MR4: parasite line 3D7 (MRA-151, deposited by D. Walliker); parasite line 7G8 (MRA-152, deposited by D. Walliker); parasite line HB3 (MRA-155, deposited by T.E. Wellems); parasite line Dd2 (MRA-156, deposited by T.E. Wellems); parasite line K1 (MRA-159, deposited by D.E. Kyle); parasite line V1/S (MRA-176, deposited by D.E. Kyle); parasite line RO-33 (MRA-200, deposited by D. Walliker, U. Certa and R. Reber-Liske); parasite line D10 (MRA-201, deposited by Y. Wu); parasite line TM90C2A (MRA-202, deposited by D.E. Kyle); parasite line TM90C6A (MRA-205, deposited by D.E. Kyle); parasite line TM91C235 (MRA-206, deposited by D.E. Kyle); parasite line WR87 (MRA-284, deposited by D.E. Kyle); parasite line D6 (MRA-285, deposited by D.E. Kyle); parasite line Malayan Camp (MRA-330, deposited by L.H. Miller and D. Baruch); parasite line Indochina I (MRA-347, deposited by W.E. Collins); parasite line Santa Lucia (MRA-362, deposited by W.E. Collins); parasite line FCC-2 (MRA-733, deposited by W. Trager); and parasite line T2-C6 (MRA-818, deposited by X. Su). Patient samples were obtained as part of ongoing studies in Senegal and Malawi described elsewhere in accordance with human subject guidelines. Additional parasites were the kind gift of: Alejandro Miguel Katzin (51, 10_54, 36_89, and 9_411); Christian Happi (APO41); Abdoulaye Djimde and Chris Plowe (PS189); Joseph Smith (A4); Karen Day (Muz51.1); Dennis Kyle and Sodsri Thaithong (TD203, TD257, TM327, TM345, GH2, and PR145); Sandra do Lago Moraes (JST); and Xin-zhuan Su

(MR24). DNA from *P. reichenowi* was kindly provided by John Barnwell.

PCR Genotyping

Genomic regions (458850-459204) surrounding the *pfcrt* (MAL7P1.27) locus were amplified by polymerase chain reaction using oligonucleotide primer sequences (CCTTGTCGACCTTAACAGATG, CTATTCCACCTACCAATATAAAAC) and the resulting DNA sample was sequenced using standard methods. In a similar manner the genomic region (754984-755584) surrounding the *dhfr* locus (PFD0830w) (oligonucleotide primer sequences: CAAGATTGATACATAAAGATAATAT, TTCTTGATAAACAACGGAACCTCCT); and the *pfmdr1* locus (PFE1150w) (oligonucleotide primer sequences: TGTTGAAAGATGGGTAAAG AGCAGAAAGAG, TACTTTCTTATTACATATGACACCACAAACA) were utilized [1].

SNP Discovery

The SNP discovery methodology was similar to that described in Volkman 2007 [2]. 1X ABI shotgun sequence was obtained for nine geographically diverse parasite isolates that were previously sequenced to 0.25X coverage, bringing total coverage to 1.25X per isolate. These nine isolates include: 7G8, Santa Lucia (El Salvador), V1/S, D10, FCC-2/Hainan, D6, RO-33, Senegal V34.04 and K1. Three of the twelve previously sequenced isolates in Volkman 2007 were excluded from additional sequencing, as they were previously found to be nearly genetically identical, suggesting possible contamination in culture [2]. Read ends with low quality (PHRED <10) bases were trimmed. Reads less than 100 bases, containing greater than 3% internal N's, or containing a mononucleotide repeat covering greater than 80% of the read were discarded. Reads were aligned to the PlasmoDB version 5 of the 3D7 genome using BLAT37 [3] requiring 95% identity, a minimum score of 100, less than 20% gaps, and coverage

of at least half of the read. Only the highest scoring alignment for each read was kept and paired reads which aligned more than 10kb apart or in the wrong orientation were discarded. The Neighborhood Quality Standard (NQS) algorithm was used to distinguish real polymorphisms from sequence errors [4]. We required the SNP to have a minimum quality score of 25, and the five base neighborhood to have a minimum score of 20. We allowed one mismatch and no indels in the neighborhood. We discarded SNPs when another read from the same sample met the NQS criteria at that position but did not have a sequence difference.

Array Development and Assessment

Based on all 111,536 discovered SNPs [2,5,6] in *P. falciparum,* and given design parameters and unique sequence constraints, we were able to design assays for 74,656 markers. Each of 74,656 SNPs is represented by a probe set of 12 to 84 probes, for a total of 4.4 million genotyping probes on the Affymetrix 49-format array. These were hybridized to 63 unique samples (totaling 81 arrays with replicates). Genotype calls were produced using the BRLMM-P algorithm, a variant of the RLMM algorithm [7], included in Affymetrix Power Tools version 1.8.5, and clustered over all 81 arrays. BRLMM-P was forced into a haploid calling mode by setting assigning all SNPs to the "Y chromosome" and setting all arrays to "male".

The array with sample TM93C1088 is eliminated immediately after clustering (arbitrarily, since the chip claiming to be TM90C6A and the chip claiming to be TM93C1088 are identical). We also remove samples CF04.010 and Senegal Th10.04, which were suspected to be multi-clonal based upon molecular barcode analysis [8]. A halofuginone-resistant version of Dd2, a human-DNA sample, and the *P. reichenowi* ancestral samples were also removed at this stage, leaving 57 unique samples (totaling 75 arrays with replicates) for analysis. We then

calculate a call rate for each SNP and removed 7,778 SNPs that had below an 80% call rate, leaving 66,878 SNPs. Since technical replicates showed 99.9% repeatability between chips, we merged replicate data for each of the 57 samples, producing a no-call when the replicates indicated discordant genotypes.

Concordance against sequencing data was calculated in both major and minor alleles for 17 sequenced reference strains [9]. The following 17 samples were compared against sequencing data for concordance: 3D7, Dd2, FCC-2, Malayan Camp, D10, K1, V1/S, RO33, D6, Senegal P31.01, Senegal P51.02, Senegal V34.04, Senegal V35.04, 7G8, A4 (subclone of IT04 [10]), Santa Lucia, and HB3. These are the 18 parasites presented in Fig 1 of Volkman 2007 [2], removing the three found to be genetically identical, and adding the two strains 3D7 and A4. A total of 18,303 SNPs lacked call overlap between array genotypes and sequencing genotypes in minor alleles and were thus removed, since concordance in both alleles could not be fully calculated. Another 30,993 SNPs were removed due to imperfect concordance, and of these discordant SNPs, most (28,789) exhibited monomorphic behavior on the array, suggesting that much of the discordance may be attributed to either a faulty assay or false discovery. The remaining 17,582 perfectly concordant SNPs constituted the high confidence set of assays used in our analyses.

Copy Number Variation (CNV) Analysis

We examined the ability to detect copy number variants (CNV) using the array by first studying a known CNV using the hybridization intensity signal of the SNP genotyping probes on the array. Kidgell, et al. (2006) [11] reported that the *pfmdr1* locus was present in 3-4 copies in the Dd2 strain relative to a collection of other strains. We compared *Z-scores* of the normalized

hybridization intensity of perfect match probes for SNPs in the neighborhood of *pfmdr1* for Dd2 and six parasites estimated by Kidgell, et al. to contain only 1 copy of the locus (3D7, 7G8, HB3, D10, D6, K1). For each SNP assay we utilized the average hybridization intensity of all perfect-match probes. Hybridization intensity values were background corrected and normalized to reduce inter-array variation artifacts. SNPs with a hybridization intensity standard deviation equal to or greater than half the magnitude of the average hybridization intensity across all arrays were excluded from analysis. Figure S11 illustrates that probes for many of the SNPs assayed within the *pfmdr1* locus exhibit notably higher hybridization intensity values in Dd2 relative to the other parasites, with 13 assays exhibiting average intensities greater than 2 standard deviations higher than observed in the other parasites.

Genome-wide Association Study (GWAS)

We performed GWAS for drug resistance to thirteen antimalarials: amodiaquine (ADQ), artemether (ARTM), artesunate (ARTN), artemisinin (ARTS), atovaquone (ATV), chloroquine (CQ), dihydroartemisinin (DHA), halofuginone (HFG), halofantrine (HFN), lumefantrine (LUM), mefloquine (MFQ), piperaquine (PIP) and quinine (QN). 50 out of 59 samples had drug phenotype data. IC50 data are shown in Table S4 and Figure S13 for these 50 parasites against the 13 drugs.

The following drugs were obtained from Sigma Aldrich: artemisinin, dihydroartemisinin, chloroquine, mefloquine, and quinine. The following were obtained from AK Scientific: artemether, artesunate, halofantrine, lumefantrine, and piperaquine. The following were obtained from USP: amodiaquine and atovaquone. Each drug was tested in triplicate for each parasite. Additionally, some parasites were tested with multiple biological replicates: 3D7 (nine biological

replicates per drug, each in triplicate), Dd2 (three replicates) and RO-33, D10, and 207-89 (two replicates).

SNPs were filtered down to a set that contained at least 5 strains with a minor allele as well as an 80% call rate under every phenotype condition. The final data set includes 7,437 SNPs. This gives us a genome-wide significance threshold of –log10(P-value) > 5.17 by Bonferroni correction for multiple testing. For binary phenotype tests (Fisher's exact test, Fisher's permuted, CMH, and HLR), we used IC50 cutoffs shown in Table S4. For tests requiring defined geographic clusters (CMH, HLR, Fisher's permuted), the three population clusters are defined by PCA, as in the LRH analysis, and the assignments are shown in Table S1.

Pointwise *P*-values were computed using PLINK [12]. Quantile-quantile plots (qq-plots) were used to examine the resulting *P*-value distributions for inflating effects due to population structure (Fig. S7). Because most of the genome is assumed to fit the null hypothesis (most of the genome should not be in association with the phenotype), significant, early deviations from this expectation may result in a high false positive rate. The null expectation is plotted as the unity diagonal line in Figure S7. Bonferroni significance is plotted as the dashed line and Benjamini-Hochberg significance is marked with the dotted line. Since most Fisher's results show evidence of inflation, we do not report these results in Figure 2 or Table 1.

Permutations of Fisher's exact test can be used to compute empirical pointwise *P*-values based on a simulated null distribution. We used PLINK to perform this permutation while respecting the phenotype frequencies present in our three predefined population clusters. The resulting *P*-value distributions (Fig. S7) do not show inflation due to population structure, however no significant hits were found for any drug.

Similarly, the Cochran-Mantel-Haenszel (CMH) test can perform population-stratified analyses for association. We used PLINK to compute *P*-values (Fig. S7), and again, we saw appropriate corrections for population structure, but no hits reached genome-wide (Bonferroni) significance.

The Efficient Mixed-Model Association (EMMA) test was specifically designed to handle quantitative trait associations to a data set with complex population structure using a linear mixed model [13]. It calculates a genotype similarity matrix instead of discrete categories and does not require *a priori* specification of population structure. The resulting *P*-value distributions demonstrate little remaining effect from population structure (Fig. S8) while retaining power to find a number of associations at genome-wide significance (Fig. S8, 2A, Table 1).

The Haplotype Likelihood Ratio (HLR) test is a multi-marker association test [14]. Unlike a standard, χ^2-based multi-marker test which looks for differences in haplotype frequencies in cases vs. controls, the HLR test specifically models the likelihood that a single haplotype rose to dominance in cases while all other haplotypes proportionally decreased. It produces a LOD score, which is the maximum likelihood estimate for the haplotype frequencies observed in cases ($O_1...O_k$), given the distribution in controls ($f_1...f_k$):

$$LOD_{ML} = \log_{10} \frac{P\left(O_1,\mathrm{K}\ ,O_k|\alpha e_j + (1-\alpha)(f_1,\mathrm{K}\ ,f_k)\right)}{P\left(O_1,\mathrm{K}\ ,O_k|f_1,\mathrm{K}\ ,f_k\right)}$$

where *j* is the haplotype on which the mutation arose, 1-α is the recombination rate, and $e_j=1$ when $i=j$ and $e_j=0$ when $i\neq j$. The test produces maximum likelihood estimates for *j* and α.

So while a χ^2-based association finds any significant differences in haplotype frequencies, the HLR test models a specific scenario that is common in rapid selective events. The HLR test does not provide significantly more power than a single marker test in regions of high LD—in extreme cases, these regions may only have two haplotypes, and a multi-marker test will have the same power as a bi-allelic SNP test. But in regions where LD is low in controls and a single, long haplotype is prevalent in cases, the HLR test is highly sensitive. The HLR test is a one-sided test and we ran separate tests for both drug resistance (called "risk") and drug sensitivity ("protect"). Results for drug sensitivity are available in Fig. S10, but are not reported generally as we are more interested in selective events for drug resistance.

We used PLINK to produce sliding window haplotypes across the genome and calculate haplotype frequencies for input to the HLR test. We produced input for all two, four and six-marker windows. The resulting LOD scores did not map well to known distributions, such as the χ^2 1-degree of freedom distribution. We instead converted the point-wise LOD scores to empirical point-wise *P*-values by performing approximately 370,000 permutations of the null model for each test condition. This allows us to calculate empirical *P*-values up to a significance of about $-\log_{10}(P\text{-value}) = 5.6$. Similar to the permuted Fisher's test, we preserved population-specific phenotype frequencies by only allowing permutations within each of our three defined populations. Resulting *P*-value distributions fit expectations well for the vast majority of test conditions (Fig. S9, S10) and the test demonstrates power to detect a number of loci at genome-wide significance (Fig. 2A, Table 1).

Supplemental Methods Figures

Figure M1. SNPs located in genes (28,576) were more likely to pass concordance filtering than intergenic SNPs (19,999).

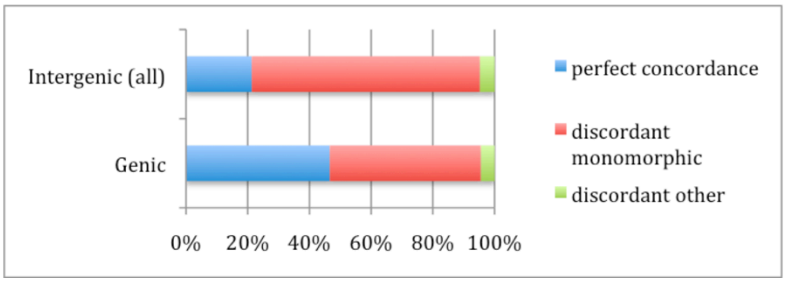


Figure M2. SNPs that were discovered from only one sequenced strain (35,727 SNPs) show a higher rate of monomorphism on the array than those with higher minor allele counts (12,848 SNPs). Any amount of this discordance may be explained by false discovery from sequencing data.

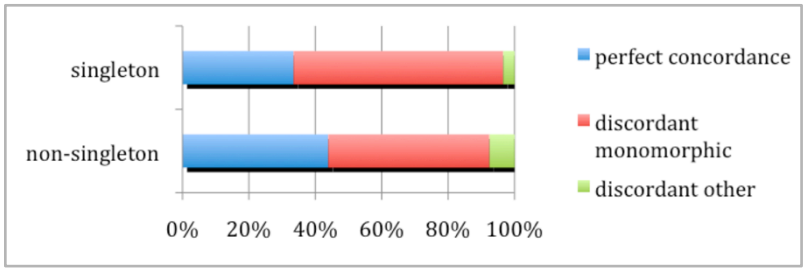


Figure M3. To show the effect of GC composition, we took 16bp of flanking sequence on each side of the SNP to construct a 3D7-based 33mer and calculated the percent GC. The window boundaries for the graph below are chosen as the octiles of the GC distribution. SNP performance appears to worsen at GC levels below 20%, which accounts for roughly half of the SNPs.

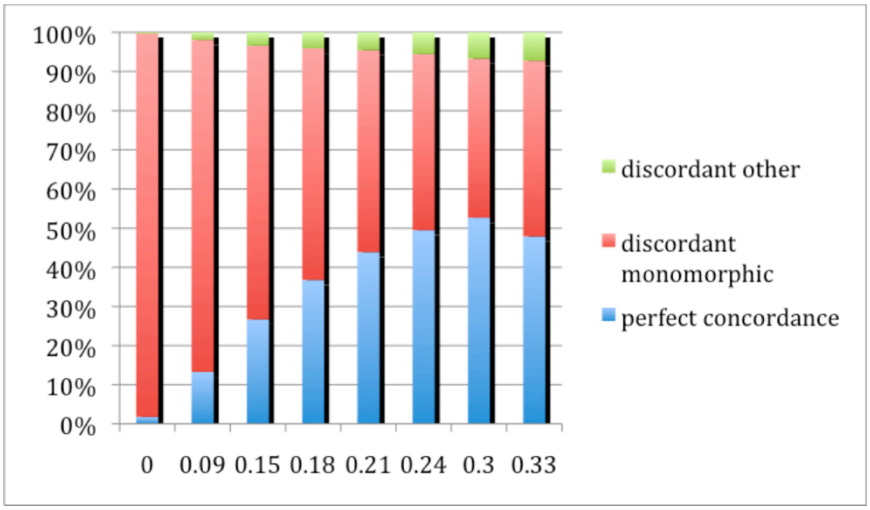


Figure M4. The effect of unique sequences in flanking regions. Although the initial design of the array excluded probes that had exact matches elsewhere in the genome, many of the remaining SNPs are in neighborhoods that contain 1 or 2 base mismatch similarity to other parts of the genome. We took 16bp of flanking sequence on each side of the SNP to construct a 3D7-based 33mer and aligned it to 3D7 using SSAHA (word length 10, step length 1, max gap 2, max insert 1, min hits 24) [15]. 28,352 SNPs aligned uniquely to their location of origin. The 20,223 SNPs that aligned in multiple locations showed a much higher rate of discordance.

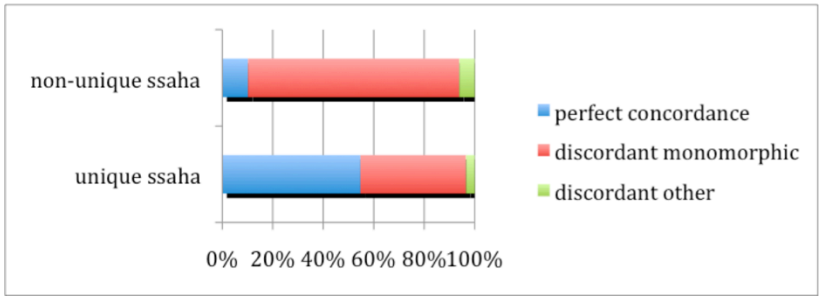


Figure M5. A histogram of marker spacing. Most markers are spaced closely with a few large gaps in coverage. The mean spacing of concordant markers is 1316bp with a median spacing of 444bp.

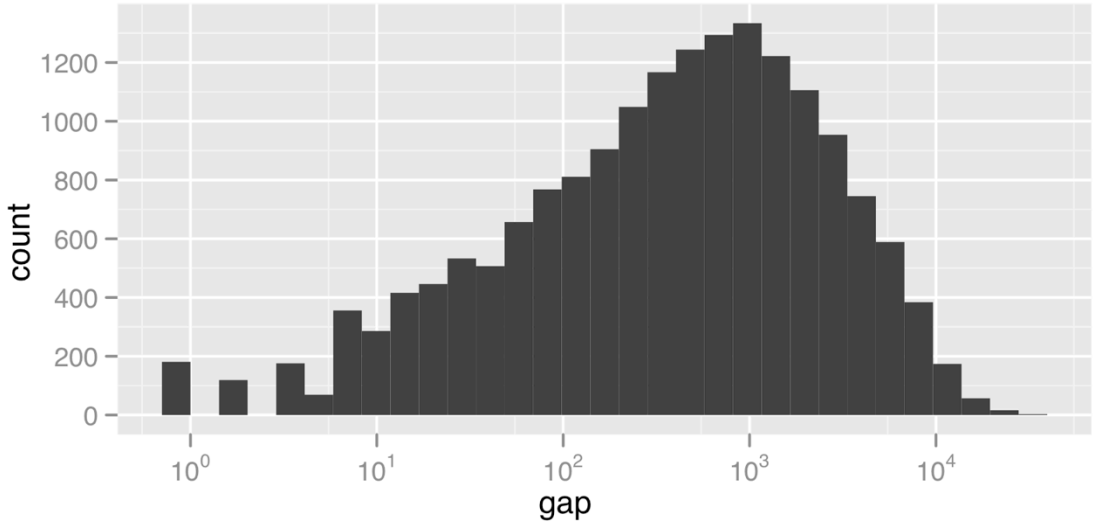


Table M1. Statistics on marker spacing (gaps) by chromosome. The table gives the number of gaps, as well as the median, mean gaps, 90% percentile (90% gap), and maximum gap length (max gap) by chromosome (1 – 14).

	number of gaps	median gap	mean gap	90% gap	max gap	sum of gaps
all	17,568	444	1,316	3,639	39,535	23,118,100
chr 1	644	150	957	2,968	18,798	616,104
chr 2	855	346	1,103	3,090	16,236	942,754
chr 3	825	485	1,282	3,300	24,076	1,057,642
chr 4	1,462	182	817	2,042	39,535	1,195,109
chr 5	977	447	1,344	3,865	17,583	1,312,630
chr 6	1,088	422	1,298	3,801	24,856	1,412,604
chr 7	1,599	225	935	2,446	19,341	1,495,349
chr 8	1,100	454	1,284	3,319	24,320	1,412,004
chr 9	1,087	557	1,415	3,820	27,661	1,538,216
chr 10	1,113	533	1,520	4,121	34,097	1,691,290
chr 11	1,357	636	1,497	4,115	21,921	2,031,460
chr 12	1,678	463	1,351	3,737	17,010	2,266,590
chr 13	1,831	669	1,562	4,333	19,966	2,860,436
chr 14	1,952	768	1,683	4,516	23,942	3,285,912

Figure M6. Final concordant marker density across all fourteen chromosomes.

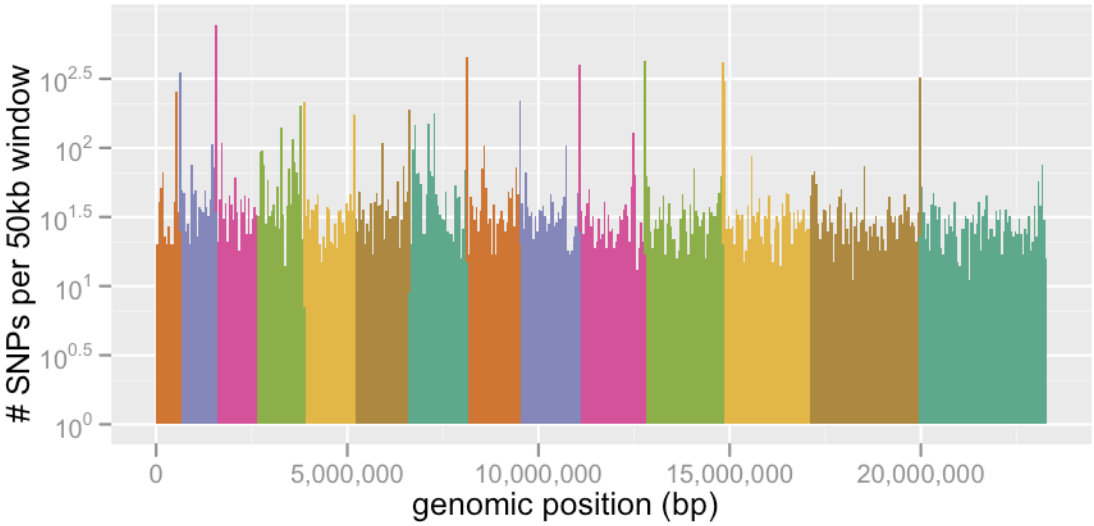
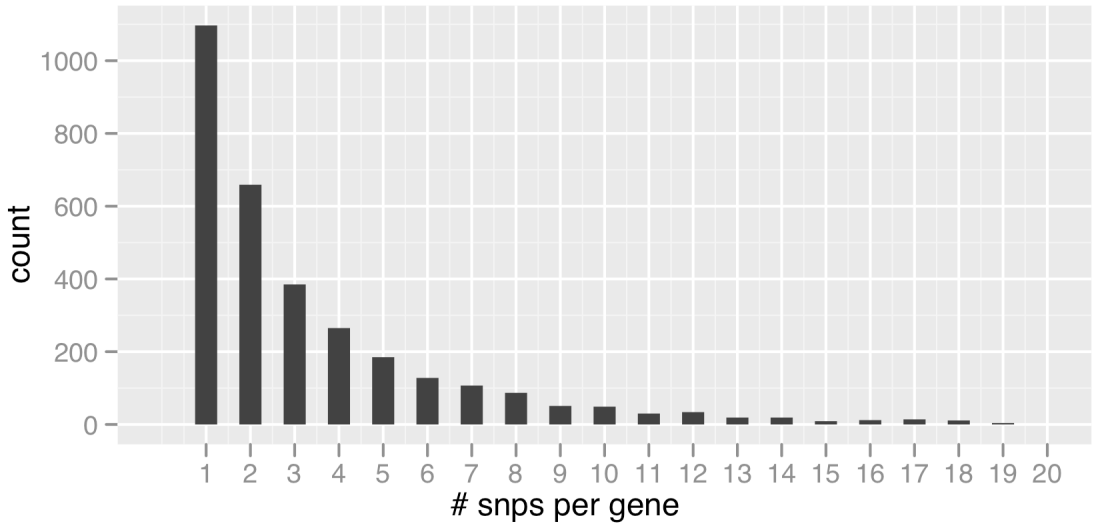


Figure M7. Distribution of markers per gene: 59% of all *P. falciparum* genes contain at least one concordant marker. Below is a plot of the number of markers per gene. Most genes that have markers have only one or two (median marker count per gene is two). Mean marker count per gene is 4.1.



Supplemental Tables

Table S1. 63 parasites used in the study with the name (parasite), geographic origin (region, country), source, and molecular barcode[8], as well as which samples were included in SNP discovery (SEQ), population characterization (POP), long-range haplotype (LRH), and GWAS analyses. For GWAS, * indicates that the sample was used, but not included in any population cluster for stratified or permuted analyses. The human control sample and the ancestral *P. reichenowi* sample were not used in any analyses reported here.

Table S1 (Continued)

Sample Information					Used in Analysis			
Parasite	*Region*	*Country*	*Source*	*Barcode*	*SEQ*	*POP*	*LRH*	*GWAS*
51	America	Brazil	Alejandro Miguel Katzin	CATTGCAGACTXCACCTTAGATTG		x		x
608	America	Brazil	Alejandro Miguel Katzin	TACCCGGGATTACAAACTAGACTT		x		x
10_54	America	Brazil	Alejandro Miguel Katzin	CACTGCAGACTXTACACTAACCTG		x		x
36_89	America	Brazil	Alejandro Miguel Katzin	TACTGCAGATCGCCCCTACGCCTG		x		x
365_89	America	Brazil	Alejandro Miguel Katzin	TGCTCCGGATTACAAACAAGACTT		x		
3D7	Europe	Netherlands	MRA-151	TACTCCGGTCCGCACCCACGATGG		x	x	*
7G8	America	Brazil	MRA-152	TACCCCAGACTXTCAATTAACCTG	x	x		x
9_411	America	Brazil	Alejandro Miguel Katzin	CACTCCAGACTGCAACTACGACTG		x		x
A4	America	Brazil	J. Smith	TATTCCGGTTCATACCCXAGATTG		x	x	
APO41	Africa	Nigeria	Christian Happi	CATTGGGGTCTACACCCAAGACTG		x	x	
CF04.008_12G	Africa	Malawi	Dan Milner	TACCCGGGACCGCCCACAAGATTG		x		x
CF04.008_1F	Africa	Malawi	Dan Milner	TATTCGGAACCGCACCCTAAATTG		x	x	x
CF04.009	Africa	Malawi	Dan Milner	TATTCCGAACCGTACCCTCGATTG		x	x	x
D10	Asia	PNG	MRA-201	CACTCCAGATTGCAACTTAGCTTG	x	x	x	x
D6	Africa	Sierra Leone	MRA-285	TACTGGAAACTGCAACCAAACTTG	x	x	x	x
Dd2	Asia	Indochina/Laos	MRA-156	CATCGCAATTCGCCCCTTAGACTG		x	x	x
FCC2	Asia	China	MRA-733	TACCCCAAATCGCACATTAACCTG	x	x	x	x
GH2	Asia	Thailand	S. Thaitong/D. Kyle	CACTGCGGTTTATCAATTAGCCTG		x	x	x
HB3	America	Honduras	MRA-155	TACTCCAGACTACACACTCACTTG		x		*
IGHCR14	Asia	India	Aditya Dash/Chetan Chitnis	TACCCCAGACCXCACACTAGACGG		x		*
Indochina_I	Asia	Indochina/Laos	MRA-347	TACTCGAGTCTACAACCACGATTG		x	x	x
JST	America	Brazil	Sandra do Lago Moraes	CACCGCGGTTTATAAACAAGATTG		x		x
K1	Asia	Thailand	MRA-159	TATTCGGATTTGTCCCTACGCCTG	x	x	x	x
M24	Africa	Kenya	X. Su	CATTGCGGTTTACCCATAAGCCTG		x	x	x
Malayan Camp	Asia	Malaysia	MRA-330	TATTCCGGATTGTCACTTAGACTG		x	x	*
Muz51.1	Asia	PNG	Karen Day	TACTCCAGATTATCACCTAGCCTG		x	x	x
PR145	Asia	Thailand	S. Thaitong/D. Kyle	CACTCCAGATCACAACCAAAACTG		x	x	x
PS189	Africa	Mali	C. Plowe/Djimde	CACTCCGGATTACAAACAAGCTTT		x	x	
RAJ116	Asia	India	Aditya Dash/Chetan Chitnis	CACTCCGAACTGCAACCACAACGG		x		*
RO33	Africa	Ghana	MRA-200	CACCCGGGATCGCAAACTAAACTT	x	x	x	x
Santa Lucia	America	El Salvador	MRA-362	CACCCGGGATTACAAACAAACCTT	x	x		*
SenP05.02	Africa	Senegal	S. Mboup	CACCCGGGATTACAAACAAGCTTT		x	x	x
SenP08.04	Africa	Senegal	S. Mboup	TACCCCGGATCGCAAACAAACTTT		x	x	x
SenP09.04	Africa	Senegal	S. Mboup	CACTCGGGTTTATACATXCAACGT		x	x	x
SenP11.02	Africa	Senegal	S. Mboup	CGCTCGAGATTACAAACTAGACTT		x	x	x
SenP19.04	Africa	Senegal	S. Mboup	TGTTCCGGTCTACAAATTAACCGT		x	x	x
SenP26.04	Africa	Senegal	S. Mboup	TATCCGAATTTATCAATACAACGT		x	x	x
SenP27.02	Africa	Senegal	S. Mboup	TACTCCGGTTTATACACTCAACGT		x	x	x
SenP31.01	Africa	Senegal	S. Mboup	TACTGCGGTCCGCAAACAAGATTG		x	x	x
SenP51.02	Africa	Senegal	S. Mboup	TATCCGGGACTGCAACTTCGACGG		x	x	x
SenP60.02	Africa	Senegal	S. Mboup	TACTCGAAACCGCAAACTAACCTT		x	x	x
SenT15.04	Africa	Senegal	S. Mboup	TGCTCCAAATCGTACCCAAGCTGT		x	x	x
SenT26.04	Africa	Senegal	S. Mboup	CGCTCGGATTTATCCCTACGCCGT		x	x	x
SenT28.04	Africa	Senegal	S. Mboup	CACCCCAGTTTATCAACAAACTGT		x	x	x
SenV34.04	Africa	Senegal	S. Mboup	TXCTCCAGATTATACCTAAACCTG	x	x	x	x
SenV35.04	Africa	Senegal	S. Mboup	CGCTCGAGTCCGTCCACACACTGT		x	x	x
SenV42.05	Africa	Senegal	S. Mboup	TGTTCCGATCTATCCACAAGACXT		x	x	x
T2_C6	Asia	Thailand	MRA-818	TACTCCGGATTATACACTAAACGT		x	x	
TD203	Asia	Thailand	S. Thaithong/D. Kyle	CGCCCCAGATCATAAATTAAACGT		x	x	x
TD257	Asia	Thailand	S. Thaithong/D. Kyle	TACTCGGGATCGTAACCACACTGT		x	x	x
TM327	Asia	Thailand	S. Thaithong/D. Kyle	TGTCCGAATCTATAAACACAACGT		x	x	x
TM345	Asia	Thailand	S. Thaithong/D. Kyle	CACCCCAGTCTATCACTACACCGT		x	x	
TM90C2A	Asia	Thailand	MRA-202	TATTCCGATTTATCAACTCACCGT		x	x	x
TM90C6A	Asia	Thailand	MRA-205	CACCCCAGTCTATCACTACACCGT		x	x	
TM91C235	Asia	Thailand	MRA-206	CGCTCCGGACTGCACCCAAGATTG		x	x	x
V1/S	Asia	Vietnam	MRA-176	TGCCCCAGATCACAACTAAGATTT	x	x	x	X
WR87	Asia	Vietnam	MRA-284	TACTGGAAATCACAACTAAGACTT		x	x	X
CF04.010	Africa	Malawi	Dan Milner	mixed				
Dd2_HFG_280	Asia	Indochina/Laos						
Human Control								
Preichenowi			John Barnwell					
SenT10.04	Africa	Senegal	S. Mboup	NACTNGGGACTATAACCAAACCTG				
TM93C1088	Asia	Thailand	MRA-207	CACCCCAGTCTATCACTACACCGT				

Table S2. Analysis of the ability of SNPs on the array to act as a proxy for or "tag" other untyped markers on the array in genome-wide association. This ability is measured using the standard correlation metric r^2. In our data set, 28% of SNPs in the Brazilian sample (which has the most LD) had a nearby SNP on the array in strong LD ($r^2 > 0.5$) with it, while in the Senegal sample the proportion was only 16%. Most of the time, therefore, we will only be able to detect association with markers that have been directly typed. The exception is strong selective sweeps, which affect many markers within a region.

Fraction of SNPs	$r^2 > 0.3$	$r^2 > 0.5$	$r^2 > 0.8$
Senegal	26%	16%	10%
Thailand	34%	24%	18%
Brazil	33%	28%	24%

Table S3. Long Range Haplotype (LRH) hits. All REHH hits with Q-value < 0.25.

chr	pos	core allele	hap len	qvalue	gene	description
2	608790	G	14428	0.2071	PFB0675w	hypothetical protein
2	617743	T	16930	0.2133	PFB0685c	acyl-CoA synthetase, PfACS9
2	623146	C	22333	0.0323	PFB0687c	RING zinc finger protein, putative
3	466483	A	35321	0.0568	PFC0460w	hypothetical protein, conserved
3	466610	A	35448	0.0133	PFC0460w	hypothetical protein, conserved
4	755220	T	15162	0.1236	PFD0830w	bifunctional dihydrofolate reductase-thymidylate synthase
4	755243	C	15139	0.2133	PFD0830w	bifunctional dihydrofolate reductase-thymidylate synthase
4	764100	G	50270	0.1528	PFD0840w	hypothetical protein, conserved
5	1042120	A	18768	0.0075	PFE1250w	acetyl-CoA synthetase, PfACS10
5	1042527	A	18361	0.0085	PFE1250w	acetyl-CoA synthetase, PfACS10
5	1056621	T	14545	0.0591		
5	1159412	C	12844	0.0085	PFE1400c	beta adaptin protein, putative
5	1159501	T	12755	0.0085	PFE1400c	beta adaptin protein, putative
5	1333609	G	5132	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333639	G	5102	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333690	T	5051	0.0664	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333703	A	5038	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333716	G	5025	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333729	T	5012	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333741	G	5000	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
5	1333790	A	4951	0.2341	PFE1640w	erythrocyte membrane protein 1 (PfEMP1), truncated
6	741192	A	55183	0.2335		
6	741293	C	55082	0.2133	PFF0855c	rifin
6	741366	A	55009	0.2133	PFF0855c	rifin
6	1025852	A	17872	0.2133	PFF1220w	hypothetical protein, conserved
6	1065237	G	33215	0.1492	PFF1280w	hypothetical protein, conserved
6	1098314	C	26112	0.0905	PFF1325c	c3h4-type ring finger protein, putative
6	1114565	C	25158	0.0905	PFF1350c	acetyl-coenzyme a synthetase
6	1114929	G	23554	0.0323	PFF1350c	acetyl-coenzyme a synthetase
6	1115373	A	23998	0.0323	PFF1350c	acetyl-coenzyme a synthetase
6	1115454	C	23938	0.0430	PFF1350c	acetyl-coenzyme a synthetase
6	1116047	G	24531	0.0244	PFF1350c	acetyl-coenzyme a synthetase
6	1116171	G	24655	0.0430	PFF1350c	acetyl-coenzyme a synthetase
6	1116315	C	24799	0.0430	PFF1350c	acetyl-coenzyme a synthetase
6	1117520	G	26004	0.0430	PFF1350c	acetyl-coenzyme a synthetase
6	1124426	C	32910	0.0535	PFF1365c	HECT-domain (ubiquitin-transferase), putative
6	1283916	G	13004	0.1978		
7	428373	C	39799	0.1269	PF07_0027	DNA-directed RNA polymerase 2 8.2 kDa polypeptide, putative
7	449953	C	31672	0.0443		
7	459787	T	21838	0.0276	MAL7P1_27	chloroquine resistance transporter
7	460216	G	21409	0.0443	MAL7P1_27	chloroquine resistance transporter
7	461218	T	20407	0.0443	MAL7P1_27	chloroquine resistance transporter
7	465826	G	39654	0.0953	PF07_0035	cg1 protein
7	465826	G	16307	0.1580	PF07_0035	cg1 protein
7	467846	G	41674	0.0130	PF07_0036	Cg6 protein
7	467846	G	14287	0.0190	PF07_0036	Cg6 protein
7	475935	T	46758	0.0664	PF07_0037	Cg2 protein
7	475935	T	26989	0.1009	PF07_0037	Cg2 protein
7	475948	A	46771	0.0664	PF07_0037	Cg2 protein
7	475948	A	26976	0.1009	PF07_0037	Cg2 protein
7	476288	G	47111	0.0591	PF07_0037	Cg2 protein

Table S3 (Continued)

7	476305	C	47128	0.0443	PF07_0037	Cg2 protein
7	476305	C	26619	0.0443	PF07_0037	Cg2 protein
7	482133	T	10978	0.0133	PF07_0038	Cg7 protein
7	482133	T	20791	0.0085	PF07_0038	Cg7 protein
7	485744	G	14589	0.0133	MAL7P1_28	ribonucleases p/mrp protein subunit, putative
7	485744	G	17180	0.0085	MAL7P1_28	ribonucleases p/mrp protein subunit, putative
7	488164	A	15424	0.0777	MAL7P1_28	ribonucleases p/mrp protein subunit, putative
7	488164	A	14760	0.0535	MAL7P1_28	ribonucleases p/mrp protein subunit, putative
7	490748	C	18008	0.0323	PF07_0040	lysophospholipase-like protein, putative
7	490748	C	14664	0.0244	PF07_0040	lysophospholipase-like protein, putative
7	490877	T	18137	0.0323	PF07_0040	lysophospholipase-like protein, putative
7	490877	T	14535	0.0133	PF07_0040	lysophospholipase-like protein, putative
7	494285	A	21431	0.0551	MAL7P1_29	hypothetical protein, conserved
7	505396	G	17088	0.2133	MAL7P1_30	hypothetical protein, conserved
7	505412	G	17104	0.2133	MAL7P1_30	hypothetical protein, conserved
7	936167	A	15183	0.0430	MAL7P1_105	hypothetical protein, conserved
7	940007	G	11940	0.2133	PF07_0085	ferrodoxin reductase-like protein
7	940111	T	12044	0.2133	PF07_0085	ferrodoxin reductase-like protein
7	940147	A	12080	0.2133	PF07_0085	ferrodoxin reductase-like protein
8	336524	T	6205	0.1930	MAL8P1_135	hypothetical membrane protein, conserved
8	452794	A	778	0.1492	PF08_0105	rifin
8	862485	A	33846	0.1239	PF08_0054	heat shock 70 kDa protein
8	866334	C	29997	0.0443	MAL8P1_64	hypothetical protein, conserved
8	1104023	T	8994	0.0591		
8	1114567	A	8029	0.0873	MAL8P1_23	ubiquitin-protein ligase 1, putative
8	1117372	G	10834	0.0905	MAL8P1_23	ubiquitin-protein ligase 1, putative
8	1118090	T	11552	0.0622	MAL8P1_23	ubiquitin-protein ligase 1, putative
8	1118190	T	11652	0.0091	MAL8P1_23	ubiquitin-protein ligase 1, putative
9	272201	C	12632	0.1259	PFI0265c	RhopH3
9	282410	G	11553	0.0932	PFI0275w	hypothetical protein, conserved
9	284833	T	12260	0.1492	PFI0280c	autophagocytosis associated protein, putative
9	284842	T	12269	0.1492	PFI0280c	autophagocytosis associated protein, putative
9	284910	G	12337	0.1492	PFI0280c	autophagocytosis associated protein, putative
10	324964	G	45666	0.2335	PF10_0078	histone deacetylase, putative
12	50106	T	288	0.0873	PFL0030c	erythrocyte membrane protein 1 (PfEMP1)
12	947550	A	63271	0.2335	PFL1130c	hypothetical protein, conserved
12	954384	C	56437	0.2250	PFL1130c	hypothetical protein, conserved
12	990296	T	43303	0.0103	PFL1170w	polyadenylate-binding protein, putative
12	1002740	T	55747	0.0568		
12	1002741	A	55748	0.0568		
14	279667	T	34730	0.0959	PF14_0074	hypothetical protein
14	960714	G	49486	0.0631	PF14_0228	hypothetical protein
14	1225984	A	8890	0.0443	PF14_0291	hypothetical protein
14	1226019	T	8925	0.0443	PF14_0291	hypothetical protein
14	1226103	C	9009	0.1528	PF14_0291	hypothetical protein
14	1226130	C	9036	0.0443	PF14_0291	hypothetical protein
14	1226242	A	9148	0.0314	PF14_0291	hypothetical protein
14	1226303	T	9209	0.0248	PF14_0291	hypothetical protein
14	1608531	A	54272	0.2133	PF14_0374	hypothetical protein
14	2812662	C	35145	0.1722	PF14_0653	hypothetical protein
14	2812679	T	35128	0.1722	PF14_0653	hypothetical protein
14	2838163	G	46456	0.2215	PF14_0660	hypothetical protein

Table S4. IC_{50} drug resistance phenotype data (nM). ND: No data.

sample	ADQ	ARTM	ARTN	ARTS	ATV	CQ	DHA	HFG	HFN	LUM	MFQ	PIP	QN
Resistance Threshold	20	5	5	10	3	50	2	1.5	5	50	20	30	100
10_54	16.48	1.551	3.079	2.002	3.335	63.57	0.7305	1.032	1.832	29.22	5.844	30.61	60.14
36_89	16.63	4.431	4.418	8.947	3.327	79.45	5.916	1.139	1.491	21.22	3.865	30.95	ND
3D7	6.8168	2.8094	3.6543	8.4817	2.6474	8.8972	2.5971	0.9846	6.6259	86.2639	20.8618	19.0066	24.5422
51	29.53	3.661	3.688	9.564	4.192	108.2	3.005	ND	2.509	20.59	7.559	41.53	125.9
608	15.16	3.089	3.07	4.351	6.978	93.25	2.784	1.155	1.198	10.09	5.83	10.47	104.9
7G8	18.37	1.464	3.1	2.988	4.161	56.18	2.789	0.7968	1.071	31.1	6.053	40.11	38.87
9_411	ND	1.158	1.66	5.99	ND	ND	0.8257	ND	1.525	5.687	0.6979	54.52	ND
CF04.008_1F	ND	3.718	3.035	6.777	0.8209	6.02	2.522	1.094	3.481	36.68	4.939	18.71	82.44
CF04.009	ND	4.1495	3.771	8.845	ND	ND	1.7316	ND	9.792	72.48	10.356	29.865	ND
D10	15.2461	3.5656	7.2633	8.7383	2.3062	11.6735	4.4617	1.5121	10.5989	92.63	25.4271	38.127	18.1534
D6	14.92	2.076	3.138	6.612	0.6697	3.611	1.333	0.4141	6.201	29.56	6.216	3.999	4.898
Dd2	10.0602	3.4412	3.5525	9.2125	1.2793	73.4253	2.3138	0.8717	4.7277	74.7528	14.9576	26.21	78.8966
FCC2	9.687	3.165	3.202	8.728	1.213	9.47	4.096	1.48	7.74	156.3	35.99	24.36	27.95
GA3	15.11	3.591	3.203	6.281	2.391	102.1	3.674	1.495	6.49	ND	27.22	43.25	31.96
GH2	23.3	6.718	5.983	16.78	2.366	98.03	5.694	1.707	16.03	346.7	17.82	48.54	149.1
HB3	8.703	1.169	3.019	3.117	1.181	9.78	0.9405	1.413	2.764	71.51	11.18	31.22	20.28
IGHCR14	4.879	0.6559	1.416	1.202	0.4847	4.967	0.382	1.451	1.725	53.57	11.51	9.564	4.99
Indochina_I	12	20.57	19.6	ND	ND	243.1	ND	0.8852	4.00E-04	6.178	3.902	28.99	153
JST	30.02	2.35	5.658	3.838	2.193	129.1	1.698	1.203	1.296	22.07	6.811	41.34	24.52
K1	16.07	1.902	3.124	3.176	2.442	86.42	3.509	1.52	1.907	30.84	13.57	35.78	77.81
M24	7.63	4.6	4.43	5.546	1.524	13.09	2.49	1.158	1.919	59.81	13.48	14.23	55.33
Malayan_Camp	7.944	2.887	1.811	1.775	1.171	7.279	2.319	1.126	1.283	ND	6.983	31.26	8.196
Muz51.1	19.64	3.29	3.18	7.449	1.939	60.44	2.271	1.13	1.696	19.5	8.501	28.23	39.73
PR145	11.11	15.79	13.66	29.7	6.59	51.24	12.02	1.106	16	140.8	53.43	31.58	149.1
RAJ116	19.23	1.075	1.805	1.367	1.315	68.59	1.035	1.546	0.0049	6.418	4.105	34.31	4.36
RO33	11.3737	2.6614	6.0825	6.9751	1.8615	11.0702	2.8762	1.6415	3.1416	69.66	8.094	34.0955	13.4901
Santa_Lucia	20.44	3.629	4.61	8.644	0.3903	11.76	3.567	1.171	0.7153	30.34	5.822	38.02	260.1
SenP05.02	16.24	5.695	3.178	11.25	2.493	ND	3.518	1.52	0.9853	30.43	6.798	29	ND
SenP08.04	17.33	4.873	4.15	11.12	1.19	14.85	3.015	0.4386	11.72	62.21	26.03	7.81	44.76
SenP09.04	4.517	5.332	4.177	15.7	0.7054	8.312	5.092	1.404	9.234	174	25.44	14.78	54.86
SenP11.02	92.05	12.75	11.15	19.99	1.881	25.83	8.258	0.7377	17.35	88.75	44.76	14.44	94.63
SenP19.04	7.157	11.66	10.98	24.04	ND	11.4	9.225	1.858	14.71	95.17	50.83	30.21	ND
SenP26.04	44.23	12.71	3.23	14.82	1.143	40.38	8.889	0.4341	20.6	57.79	84.86	5.288	ND
SenP27.02	9.269	3.232	3.635	8.222	ND	9.813	2.664	1.258	1.83	31.2	5.972	21.2	ND
SenP31.01	6.961	1.991	1.999	6.131	0.4997	8.854	2.966	1.214	3.177	78.89	17.03	22.56	ND
SenP51.02	29.62	5.041	4.748	8.875	0.3225	62.05	3.398	1.627	3.319	48.55	8.601	34.89	50.08
SenP60.02	15.8	4.084	4.978	9.62	0.9355	99.93	3.854	1.255	1.616	21.99	11.93	21.19	47.78
SenT15.04	ND	3.595	3.016	6.9315	ND	ND	3.3665	ND	5.262	51.32	3.948	49.53	ND
SenT26.04	16.05	3.149	3.116	4.827	1.19	79.98	2.991	1.43	2.905	30.92	7.877	24.72	121.1
SenT28.04	11.9	3.744	3.139	7.846	1.1	51.01	2.343	1.461	7.5	62.69	32.64	15.63	139.6
SenV34.04	21.39	5.1	4.645	6.781	0.5187	ND	3.024	1.627	2.37	30.34	4.018	28.48	ND
SenV35.04	5.567	3.078	3.68	4.706	0.8165	9.364	5.374	1.359	2.761	45.52	20.6	26.28	ND
SenV42.05	6.456	1.06	1.565	1.717	0.7363	9.594	1.731	1.016	4.476	13.6	10.04	4.485	ND
TD203	15.63	8.15	4.736	20.56	8.49	52.86	8.507	1.071	7.938	156.2	30.32	39.2	67.53
TD257	11.68	11.88	8.532	24.53	2.215	55.41	11.63	1.539	15.96	227.6	50.42	35.53	253.5
TM327	9.891	6.032	1.7	8.862	1.146	41.12	3.851	2.022	15.49	249.3	41.25	50.89	70.72
TM90C2A	18.85	16.17	10.31	28.32	3.704	58.94	5.376	1.086	12.51	264.5	28.04	30.04	209.1
TM91C235	ND	12.43	12.46	17.66	ND	ND	7.409	ND	70.65	154.5	57.96	12.69	ND
V1/S	17.98	1.625	3.459	6.172	4.02	155.5	1.035	1.538	1.739	26.4	12.84	38.58	224.3

Table S5. Parasites used in the GWAS indicating their nucleotide and amino acid sequence for various positions (indicated by number) in the *dhfr*, *pfcrt*, and *pfmdr1* gene loci.

	dhfr *Amino Acid Sequence*						***pfcrt*** *Amino Acid Sequence*							***pfmdr1*** *Amino Acid Sequence*				
Parasite	16	51	59	108	164	506	72	74	75	76	326	356	371	86	184	1034	1042	1246
51	A	I	C	N	I	Y	S	M	N	T	N	L	R	N	F	C	D	Y
10_54	A	I	C	N	I	Y	S	M	N	T	N	L	R	N	F	C	D	Y
36_89	A	I	C	N	I	Y	S	M	N	T	N	L	R	N	F	C	D	Y
3D7	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
7G8	A	I	C	N	I	Y	S	M	N	T	N	L	R	N	F	C	D	Y
9_411	A	I	C	N	I	Y	S	M	N	T	N	L	R	N	F	C	D	Y
A4	A	I	C	T	I	Y	C	I	E	T	N	I	I	Y	Y	S	N	D
APO41	A	I	R	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
CF04.008_12G	A	N	R	N	I	Y	C	M	N	A	N	I	I	Y	Y	S	N	D
CF04.008_1F	A	I	R	N	I	Y	C	M	N	A	N	I	R	Y	Y	S	N	D
CF04.009	A	I	R	N	I	Y	C	M	N	A	N	I	R	N	F	S	N	D
D10	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
D6	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
Dd2	A	I	R	N	I	Y	C	I	E	T	N	T	I	Y	Y	S	N	D
FCC2	A	N	C	N	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
GH2	A	I	R	N	I	Y	C	I	E	T	N	I	I	N	Y	S	N	D
HB3	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	D	D
Indochina_I	A	I	R	N	I	Y	C	I	E	T	N	I	I	N	F	C	D	Y
JST	A	N	C	N	I	Y	S	M	N	T	N	L	R	N	Y	C	D	Y
K1	A	N	R	N	I	Y	C	I	E	T	N	I	I	Y	Y	S	N	D
M24	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
Malayan Camp	A	N	C	N	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
Muz51.1	A	N	C	S	I	Y	S	M	N	T	N	L	R	Y	Y	S	N	D
PR145	A	I	R	N	I	Y	G	I	E	T	N	I	I	N	F	S	N	D
PS189	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
RO33	A	N	R	N	I	Y	C	M	N	K	N	I	R	Y	F	S	N	D
Santa Lucia	A	N	C	S	I	Y	C	M	N	T	N	I	R	N	F	C	D	D
SenP05.02	A	N	C	N	I	Y	C	I	E	T	N	I	I	Y	F	S	N	D
SenP08.04	A	I	R	N	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
SenP09.04	A	I	R	N	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
SenP11.02	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
SenP19.04	A	N	R	N	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D

Table S5 (Continued)

SenP26.04	A	N	R	N	I	Y	C	I	E	T	N	I	I	Y	F	S	N	D
SenP27.02	A	I	R	N	I	Y	C	M	N	K	N	I	R	Y	F	S	N	D
SenP31.01	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
SenP51.02	A	I	R	N	I	Y	C	I	E	T	N	I	I	Y	F	S	N	D
SenP60.02	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
SenT15.04	A	I	R	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
SenT26.04	A	T	R	N	I	Y	C	I	E	T	N	I	I	Y	F	S	N	D
SenT28.04	A		R	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
SenV34.04	A	N	C	S	I	Y	C	I	E	T	N	I	I	Y	F	S	N	D
SenV35.04	A	N	C	S	I	Y	C	M	N	K	N	I	R	N	F	S	N	D
SenV42.05	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	Y	S	N	D
T2_C6	A	N	R	S	I	Y	C	M	N	K	N	I	R	N	Y	S	N	D
TD203	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
TD257	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
TM327	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
TM345	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
TM90C2A	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
TM90C6A	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	Y	S	N	D
TM91C235	A	I	C	N	I	Y	C	I	E	T	N	I	I	N	F	S	N	D
V1/S	A	I	C	N	I	Y	C	I	E	T	N	I	I	Y	Y	S	N	D

Table S6. *PF10_0355* copy number summary for 38 parasites tested by qPCR using the Delta Delta Ct method. Copy number (CN) was compared to the reference locus *PF07_0076* and 3D7 was used as a reference strain. A cut-off of 1.4 was used to define *PF10_0355* copy number greater than 1; parasites with greater than 1 copy of *PF10_0355* are shaded. Parasites are ranked by Halofantrine (HFN) IC_{50}: HFN-sensitive parasites are indicated by an S and HFN-resistant parasites are indicated by an R.

Parasite	CN	HFN
Indochina_I	0.92	S
RAJ116	0.92	S
Santa_Lucia	1.03	S
SenP05.02	0.98	S
7G8	0.74	S
Malayan_Camp	0.94	S
JST	1.06	S
36_89	1.07	S
SenP60.02	1.19	S
Muz51.1	0.64	S
IGHCR14	1.63	S
V1/S	0.89	S
10_54	0.99	S
K1	0.87	S
M24	0.76	S
SenV34.04	4.92	S
51	0.77	S
SenV35.04	0.86	S
HB3	0.88	S
RO33	1.26	S
SenP31.01	0.77	S
SenP51.02	1.09	S
CF04.008_1F	0.95	S
SenV42.05	1.26	S
Dd2	1.07	S
SenT15.04	1.71	R
3D7	1	R
SenT28.04	0.73	R
FCC2	0.76	R
TD203	0.9	R
D10	1.06	R
SenP08.04	0.95	R
TM90C2A	1.43	R
SenP19.04	1.06	R
PR145	0.94	R
GH2	1.71	R
SenP11.02	7.14	R
SenP26.04	1.68	R

Table S7. Annotation and GeneID Information for identified genes in Figure 1B.

GeneID	π	F_{ST}	Annotation	Category	Tag
MAL8P1_23	1.54E-04	0.646	ubiquitin-protein ligase 1, putative	enzymes, ACS and transporters	UBQ Ligase
PF13_0201	6.39E-03	0.216	thrombospondin-related anonymous protein, TRAP	other	TRAP
PFA0650w	4.37E-03	0.323	surface-associated interspersed gene pseudogene, (SURFIN) pseudogene	antigens, var, rifin, stevor, surfin	SURFIN
PF08_0105	6.10E-03	0.204	rifin	antigens, var, rifin, stevor, surfin	Rifin
PFB0960c	4.31E-03	0.036	P. falciparum Maurer's Cleft 2 transmembrane domain protein 2.1, PfMC-2TM_2.1	other	Mauer's Cleft
MAL7P1_27	6.36E-04	0.387	chloroquine resistance transporter	enzymes, ACS and transporters	PFCRT
PF10_0345	6.52E-03	0.240	merozoite surface protein 3	antigens, var, rifin, stevor, surfin	MSP3
PFI1475w	1.95E-03	0.221	merozoite surface protein 1, precursor	antigens, var, rifin, stevor, surfin	MSP1
PFB0972w	9.90E-03	0.077	hypothetical protein	other	*
PFL0030c	7.95E-03	0.050	erythrocyte membrane protein 1 (PfEMP1)	antigens, var, rifin, stevor, surfin	Var2CSA
PFD0830w	5.96E-04	0.459	bifunctional dihydrofolate reductase-thymidylate synthase	enzymes, ACS and transporters	DHFR
PF11_0344	6.46E-03	0.074	apical membrane antigen 1, AMA1	antigens, var, rifin, stevor, surfin	AMA1
PF10_0051	5.32E-03	0.215	ADP/ATP carrier protein, putative	enzymes, ACS and transporters	ADP/ATP Carrier
PFB0685c	5.75E-04	0.497	acyl-CoA synthetase, PfACS9	enzymes, ACS and transporters	ACS9
PFF1350c	2.00E-03	0.584	acetyl-coenzyme a synthetase	enzymes, ACS and transporters	ACS
PFE1250w	1.66E-03	0.602	acyl-CoA synthetase, PfACS10	enzymes, ACS and transporters	ACS10

Supplemental Figures

Figure S1. Principal components analysis of population structure within A) Africa B) the Americas, and C) Asia. Plots of the first two principal components using Eigenstrat [16] using the Affymetrix array. Each solid circle represents an individual, and the color is assigned according to the reported origin.

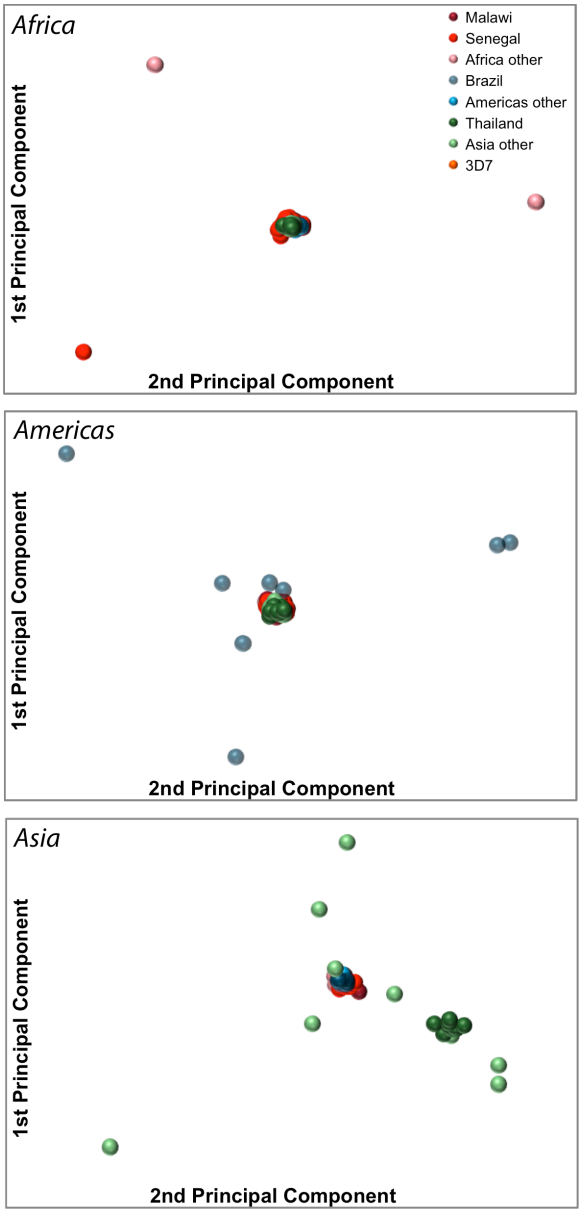


Figure S2: Linkage disequilibrium (LD), measured by r^2, for each of the three population samples (Senegal, Thailand, Brazil). Plotted are r^2 for linked markers (red lines) and for unlinked markers (blue lines), as well as the level of background LD expected because of small sample size (green lines).

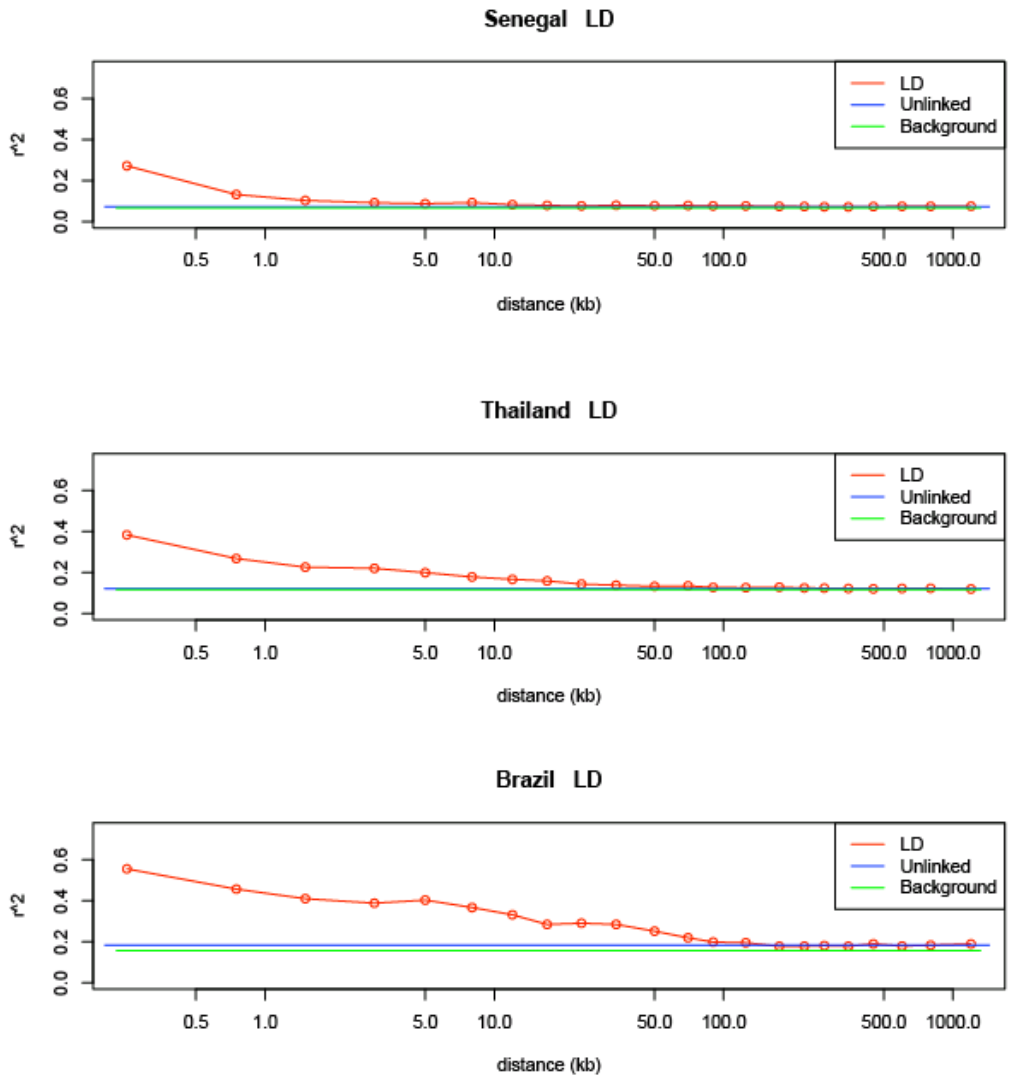


Figure S3. Genes were classified by gene ontology (GO) functional categories and stratified by level of nucleotide diversity (π) as estimated by *Z*-scores. Select categories (highest five and lowest five categories along with categories in between that differ by incremental *Z*-scores) are shown. The majority of genes in GO categories for molecules found at the cell membrane have high levels of nucleotide diversity, while most of the genes classified into GO categories for conserved molecules lack nucleotide diversity.

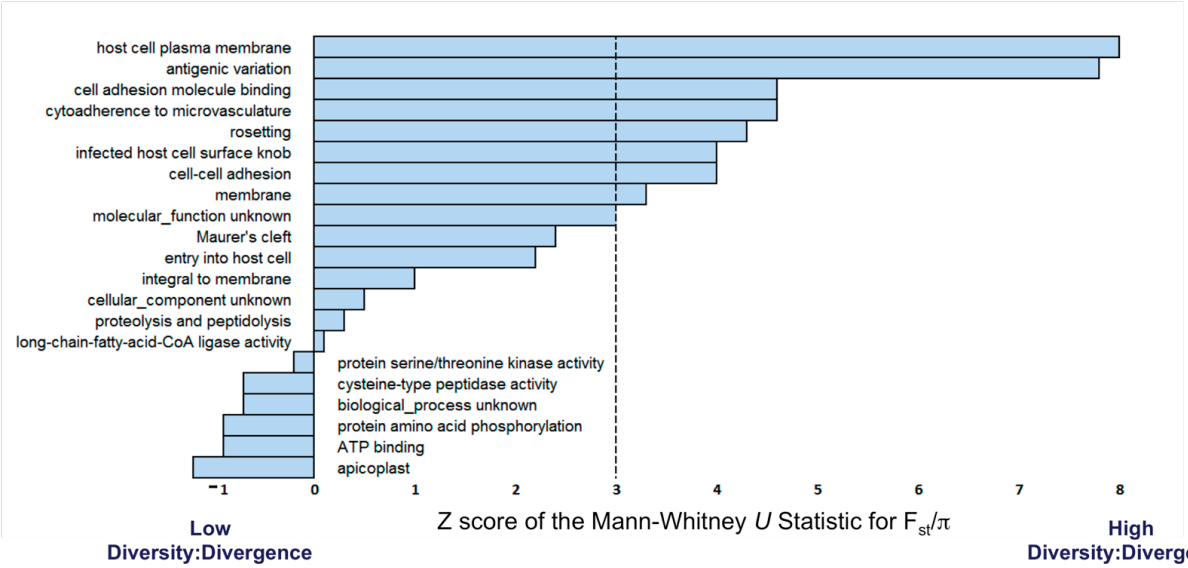


Figure S4. SNP diversity and divergence by translation consequence. Diversity at assayed SNPs (SNP π) and Divergence between different populations as assayed by F_{ST}, for different classes of SNPs: intergenic (4,263 SNPs), intronic (584 SNPs), synonymous (3,957 SNPs), and nonsynonymous (8,778 SNPs). Intronic SNPs have the widest error bars due to their relative sparseness on the array. Non-synonymous SNPs are generally among the least diverse and most differentiated class of SNPs.

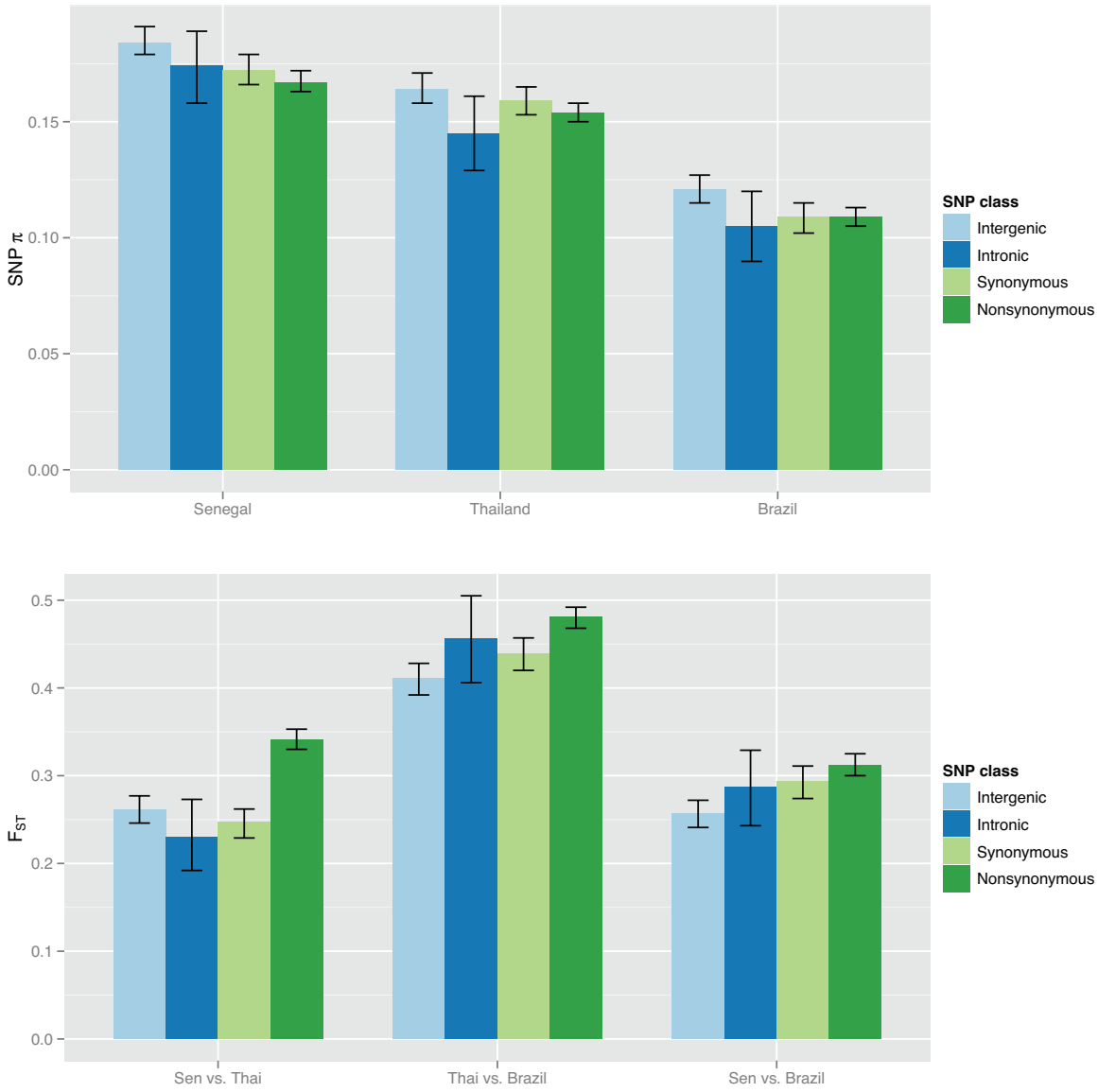


Figure S5. Relative extended haplotype homozygosity (REHH) scores. Relative extended haplotype homozygosity (REHH) scores prior to any normalization, plotted for each core allele, (A) indexed by chromosome and position, and colored by chromosome, and (B) as a function of core allele frequency.

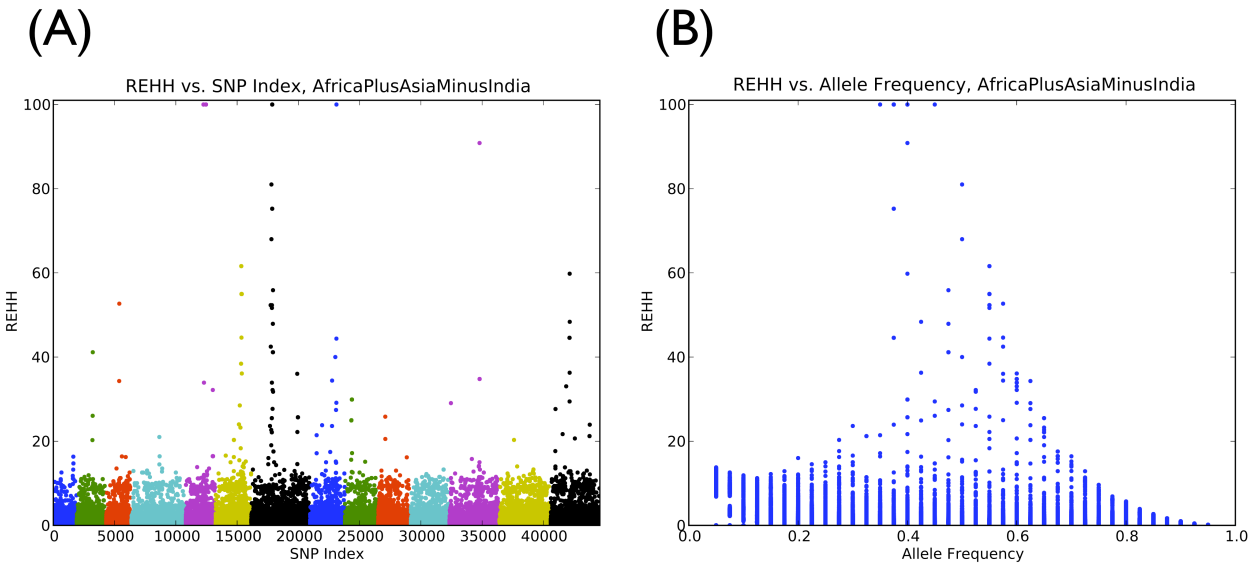


Figure S6. Long-range haplotype (LRH) analysis yields genome-wide significant candidates for recent positive selection. For each core allele, we calculated relative extended haplotype homozygosity (REHH), and from the set of all REHH scores we calculated a corresponding distribution of *Q*-values. We plotted -$\log_{10}$(*Q*-value), for all *Q*-values <1, for each core allele, indexed by chromosome and position, and colored by chromosome. The red dotted line corresponds to the typical *Q*-value significance threshold of 0.05. Gene annotations from PlasmoDB.org for some significant scores are labeled. For comparison, the well-known sweeps around drug resistance loci *pfcrt* and *dhfr* are labeled. This data is also shown in tabular form in Table S3.

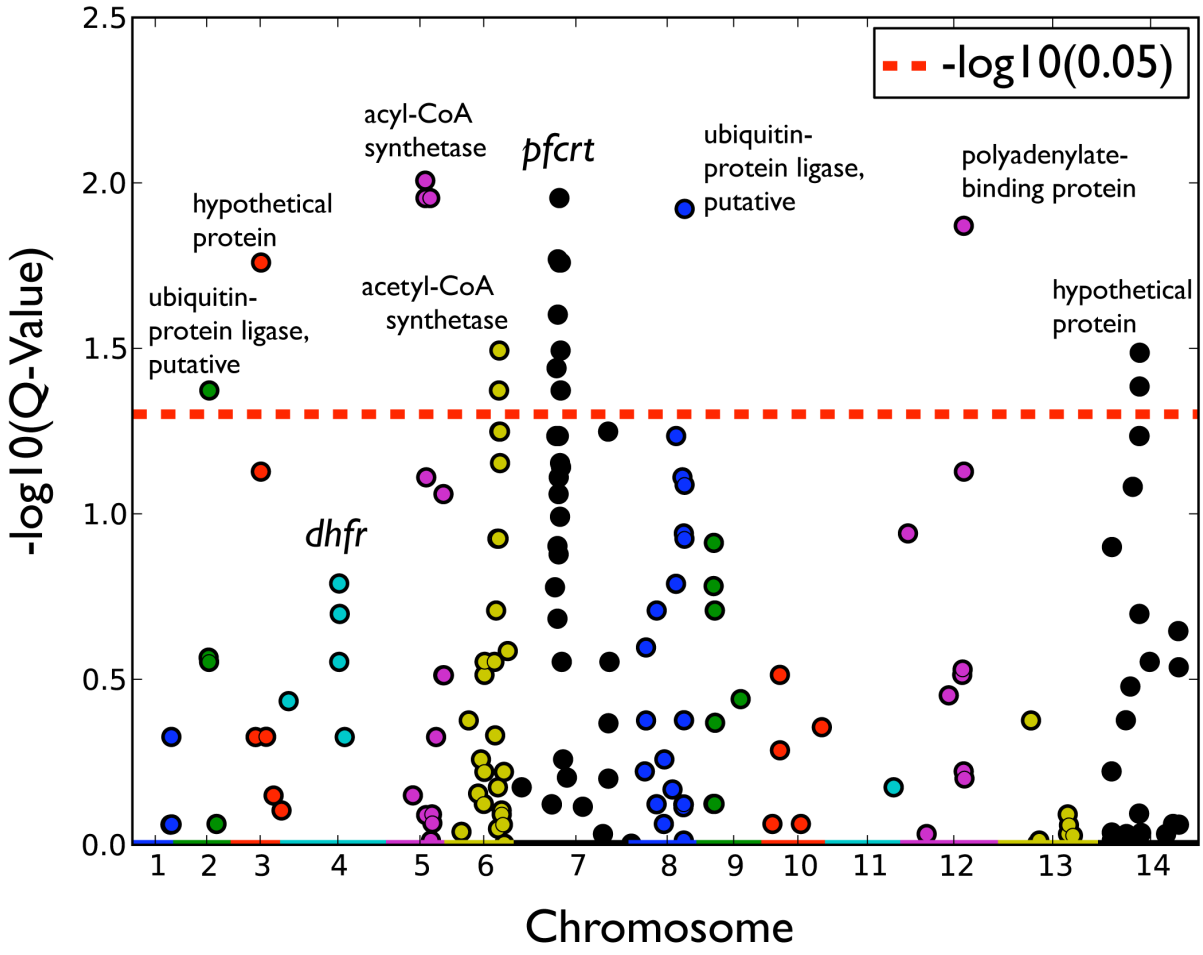


Figure S7. GWAS *P*-value distributions for Fisher's exact test, permuted Fisher's exact test, and Cochran-Mantel-Haenszel (CMH) tests. Quantile-quantile plots (qq-plots) show log *P*-values for every SNP on the y-axis against the null expectation on the x-axis. Fisher's exact test results generally show *P*-value inflation due to confounding effects from population structure for many drugs ("Fish"). As such, no results from this test are reported. To account for population structure, permutations of the null distribution were performed while preserving phenotypic associations to three predefined population clusters ("Fishp"). CMH also performs a stratified association test given predefined population clusters ("CMH"). The permuted Fisher's test and CMH test results show appropriate correction for population structure, but show no hits at genome-wide significance to report.

Figure S7 (Contined)

Figure S8. GWAS results for the Efficient Mixed-Model Association (EMMA) test. QQ-plots show little to no confounding effect from population structure, with the possible exception of artesunate (ARTN). The significant ARTN result is not reported in Table 1 or Figure 2 for this reason. Manhattan plots depict the genomic location of significant hits, also reported in Table 1 and Figure 2.

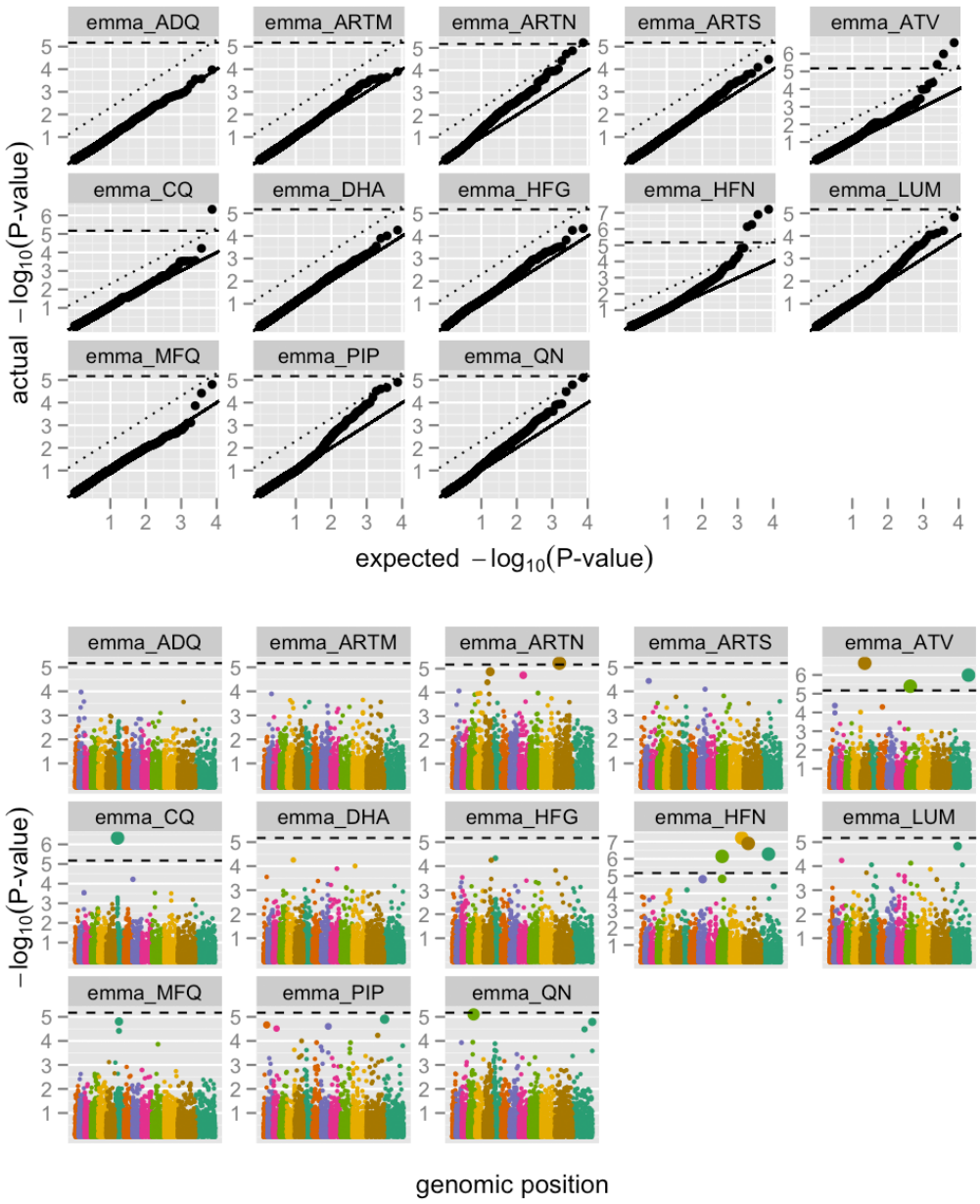


Figure S9. GWAS *P*-value distributions for the Haplotype Likelihood Ratio (HLR) tests for association for drug resistance. Population-sensitive permutations of the null model were used to calculate *P*-values from LOD scores. Final distributions of *P*-values show little to no confounding effect from population structure for most tests. Exceptions include the 6-SNP artemether (HLR_risk_6_ARTM) test and the 4-SNP amodiaquine (HLR_risk_4_ADQ) test--these results are not reported in Table 1 or Figure 2. Manhattan plots for other tests that reached genome-wide significance are in Figure 2A.

Figure S9 (Continued)

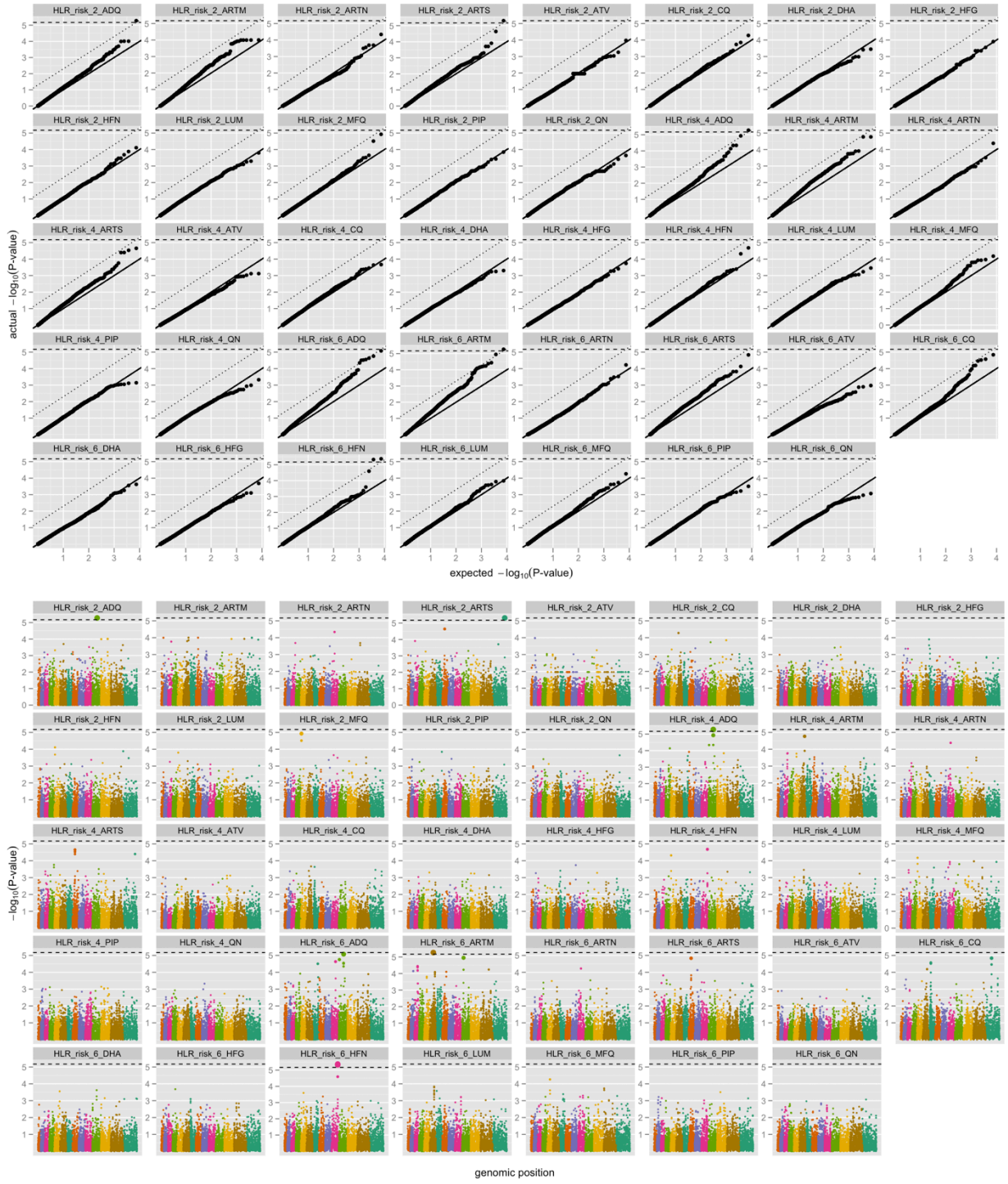


Figure S10. GWAS *P*-value distributions for Haplotype Likelihood Ratio (HLR) tests for association for drug sensitivity. Population-sensitive permutations of the null model were used to calculate *P*-values from LOD scores. Final distributions of *P*-values show little to no confounding effect from population structure. Genome-wide significant hits include piperaquine (HLR_protect_4_PIP) on a haplotype that spans *PF07_0126*, *PF07_0127* and *MAL7P1_167* and amodiaquine (HLR_protect_4_ADQ) on a haplotype in *PFL1800w*. A chloroquine hit on *pfcrt* just misses genome-wide significance. These results are not reported in Table 1.

Figure S10 (Continued)

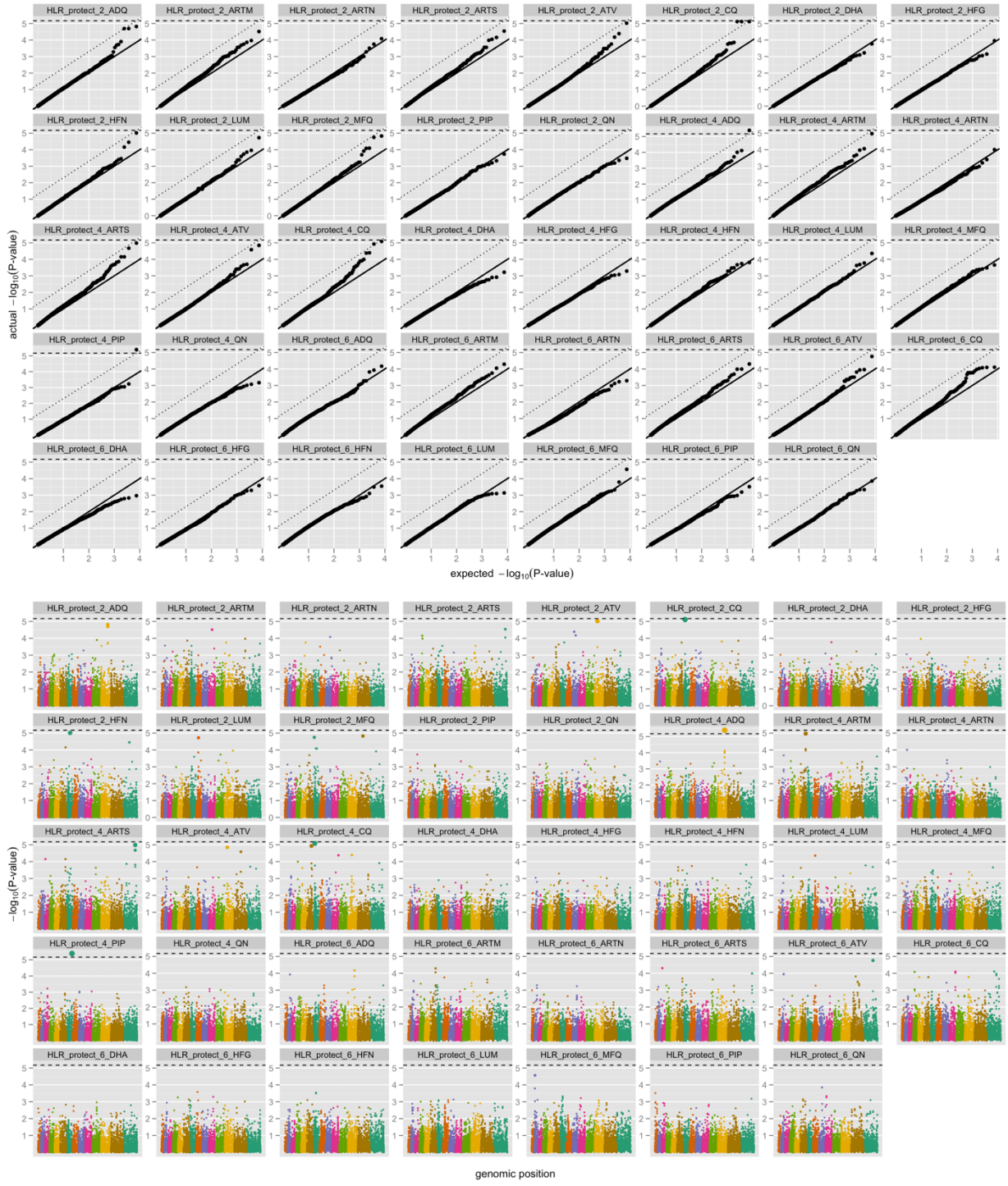


Figure S11. Intensity *Z*-score for the Affymetrix array across chromosome 5. The results illustrate that probes for many of the SNPs assayed within the *pfmdr1* (888-988 k) locus exhibit notably higher hybridization intensity values in Dd2 relative to the other parasites, with 13 assays exhibiting average intensities greater than 2 standard deviations higher than observed in other strains. This is consistent with the copy number variation reported in the *pfmdr1* locus, with 3–4 copies present in the Dd2 strain relative to a collection of other strains.

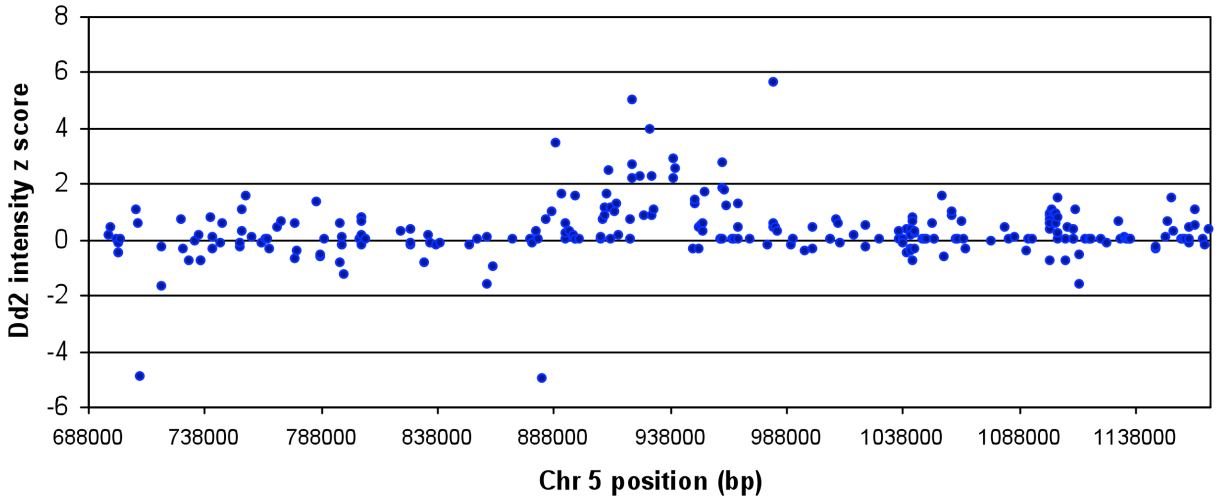
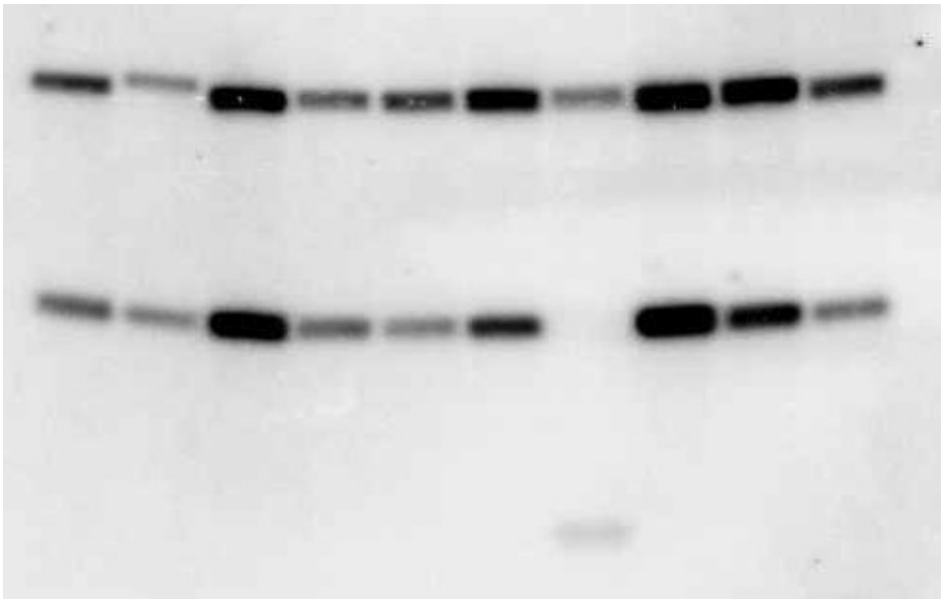


Figure S12. *PF10_0355* copy number variation measured by Southern blotting. Select parasite isolates were digested with AflIII, EcoRV and XbaI and fragments were detected using probes to portions of the *PF10_0355* and *dhps* genes. Primers used for making probes were: *dhps* F: 5′-GTG ATT GTG TGG ATC AGA AGA TGA ATA ATC-3′; R: 5′-GGA TTA GGT ATA ACA AAA GGA CCA GAG G-3′; *PF10_0355* F: 5′-GGG GAA AGC ATA TAA TAA TAC TAT AGA TGC-3′; R: 5′-CTT GGA GGA ACA AGA ACC CCC TTA TTA TCA-3′ Radioactivity was measured using a phosphorimager plate and quantified using Quantity One software (version 4.6.5). Halofantrine (HFN) response is listed as sensitive (S) or resistant (R) for each strain.



	V34.04	TM90C6A	TM90C2A	TM327	P26.04	P19.04	P11.02	GH2	D10	3D7
PF10_0355 copy number	1.54	**2.26**	**2.43**	**2.02**	1.21	1.18	0.53	**2.38**	1.06	1.00
HFN response	S	R	R	R	R	R	R	R	R	R

Figure S13. Drug resistance phenotype classification for sweep and GWAS analyses. IC_{50} data were collected for thirteen anti-malarial drugs against all genotyped parasite lines. Quantitative IC_{50}s were converted into binary "sensitive" and "resistant" phenotypes at the cutoffs shown (see also Table S4). These binary phenotypes were only used for the Haplotype Likelihood Ratio (HLR) test. Drug abbreviations: amodiaquine (ADQ), artemether (ARTM), artesunate (ARTN), artemisinin (ARTS), atovaquone (ATV), chloroquine (CQ), dihydroartemisinin (DHA), halofuginone (HFG), halofantrine (HFN), lumefantrine (LUM), mefloquine (MFQ), piperaquine (PIP) and quinine (QN).

References

1. Duraisingh MT, Jones P, Sambou I, von Seidlein L, Pinder M, et al. (2000) The tyrosine-86 allele of the pfmdr1 gene of Plasmodium falciparum is associated with increased sensitivity to the anti- malarials mefloquine and artemisinin. *Mol Biochem Parasitol* 108: 13-23

2. Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, et al. (2007) A genome-wide map of diversity in Plasmodium falciparum. *Nat Genet* 39: 113-119

3. Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12: 656-664. 4. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, et al. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513-516

5. Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, et al. (2007) Genome variation and evolution of the malaria parasite Plasmodium falciparum. *Nat Genet* 39: 120-125

6. Ribacke U, Mok BW, Wirta V, Normark J, Lundeberg J, et al. (2007) Genome wide gene amplifications and deletions in Plasmodium falciparum. *Mol Biochem Parasitol* 155: 33-44

7. Rabbee N, Speed TP (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22: 7-12

8. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, et al. (2008) A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. *Malar J* 7: 223

9. Neafsey DE, Schaffner SF, Volkman SK, Park DJ, Montgomery P, et al. (2008) Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. *Genome Bio*. pp. R171

10. Roberts DJ, Craig AG, Berendt AR, Pinches R, Nash G, et al. (1992) Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* 357: 689-692

11. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, et al. (2006) A systematic map of genetic variation in Plasmodium falciparum. *PLoS Pathog* 2: e57

12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575

13. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709-1723

14. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438: 803-819

15. Ning Z, Cox AJ, Mullikin JC (2001) SSAHA: a fast search method for large DNA databases. *Genome Res* 11: 1725-1729

16. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2: e190

APPENDIX B

Supplemental Methods and Figures for Chapter Two:

Van Tyne et al. (in prep). Modulation of *PF10_0355* alters *Plasmodium falciparum* response to antimalarial drugs

Supplemental Methods

PF10_0355 Immunoprecipitation

Sorbitol-synchronized schizont-stage Dd2 0355 DR parasitized red blood cells were lysed with 0.05% Saponin containing Complete Protease Inhibitor (Roche Applied Sciences, Indianapolis, IN) and then washed with 1xPBS to separate parasites from lysed red blood cells. Parasites were then lysed with a buffer containing 50mM TrisHCl pH 7.4, 100mM NaCl, 0.5% NP40, and Complete Protease Inhibitor (Roche Applied Sciences, Indianapolis, IN). Parasites were incubated in lysis buffer on ice and were vortex periodically for 10 minutes, followed by filtration through a 0.22µm filter. Anti-HA affinity matrix (clone 3F10; Roche Applied Science, Indianapolis, IN) was pre-washed with lysis buffer, and then filtered lysate was added and incubated with the affinity matrix for 1 hour at 4 degrees with gentle shaking. Affinity matrix was then spun down and supernatant (flow-through or "unbound" protein) was removed. Affinity matrix (with bound proteins) was washed 4 times with lysis buffer and was then mixed with 2x Laemmli sample loading buffer (Bio-Rad, Hercules, CA) containing beta-mercaptoethanol, boiled and analyzed by SDS-PAGE with both Coomassie blue staining (Bio-Rad, Hercules, CA) and Silver staining using a SilverQuest kit (Invitrogen, Grand Island, NY).

Protein Identification

Individual protein bands were excised and sent for analysis at the Taplin Mass Spectrometry facility at Harvard Medical School. Gel pieces were subjected to a modified in-gel trypsin digestion procedure [1]. Gel pieces were then washed and dehydrated with acetonitrile for 10 minutes, followed by removal of acetonitrile. Pieces were then completely dried in a speed-vac. Rehydration of the gel pieces was with 50 mM ammonium bicarbonate solution containing 12.5 ng/µl modified sequencing-grade trypsin (Promega, Madison, WI) at 4ºC. After

45 min., the excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate solution to just cover the gel pieces. Samples were then placed in a 37ºC room overnight. Peptides were later extracted by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile and 1% formic acid. The extracts were then dried in a speed-vac (~1 hr), and samples were then stored at 4ºC until analysis.

On the day of analysis, the samples were reconstituted in 5-10µl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 5µm C18 spherical silica beads into a fused silica capillary (125µm inner diameter x ~20cm length) with a flame-drawn tip [2]. After equilibrating the column, each sample was loaded via a Famos auto sampler (LC Packings, San Francisco CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted they were subjected to electrospray ionization and then entered into an LTQ Velos ion-trap mass spectrometer (ThermoFisher, San Jose, CA). Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching protein databases with the acquired fragmentation pattern by the Sequest software program (ThermoFisher, San Jose, CA) [3]. Spectral matches were manually examined and multiple identified peptides per protein were required.

Biacore drug binding

Recombinant protein was generated for full-length PF10_0355 and PF10_0348 proteins, as well as the DBL domain of PF10_0355. Drug binding analysis was conducted on a Biacore 3000 using CM5 chips (GE Healthcare, Waukesha, WI). Proteins were immobilized on chips

using amine coupling chemistry [4], and drug binding was assessed using a running buffer containing 5% DMSO and 0.05% P20 surfactant in 1xPBS. Data were analyzed using BIAevaluation software (GE Healthcare, Waukesha, WI), Microsoft Excel (Redmond, WA) and GraphPad Prism v5.0d (San Diego, CA). Response units (RU) were solvent corrected and normalized to a blank flow-cell that was analyzed in parallel, and K_d values were calculated using a non-linear fit model for one-site total binding.

Supplemental Figures

Figure B1. Pull-down of PF10_0355 and identification of interacting partners. A) Western blot showing fractions of lysate, flow-through ("Unbound"), wash and protein remaining on anti-HA affinity matrix ("Beads") following pull-down procedure. Anti-HA antibody staining is shown in green and anti-LDH antibody staining is shown in red. B) Silver-stained gel of protein remaining on anti-HA affinity matrix ("Beads") following immunoprecipitation. Arrows point to bands that were excised and sent for mass spectrometry analysis. Protein bands that were identified by at least five different peptide fragments are labeled with their protein matches and likely protein domain sizes.

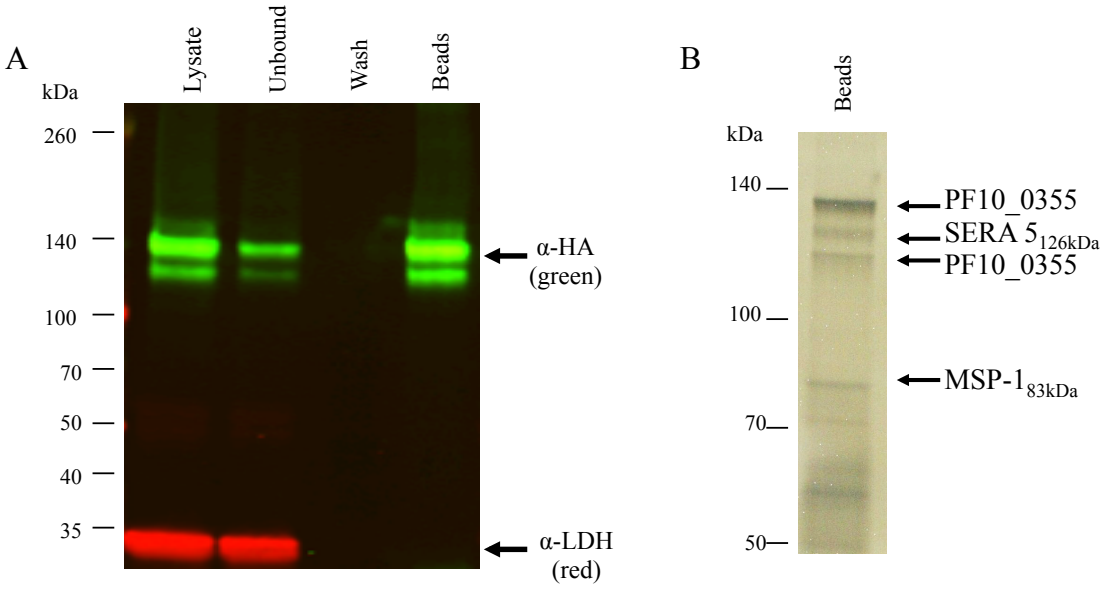
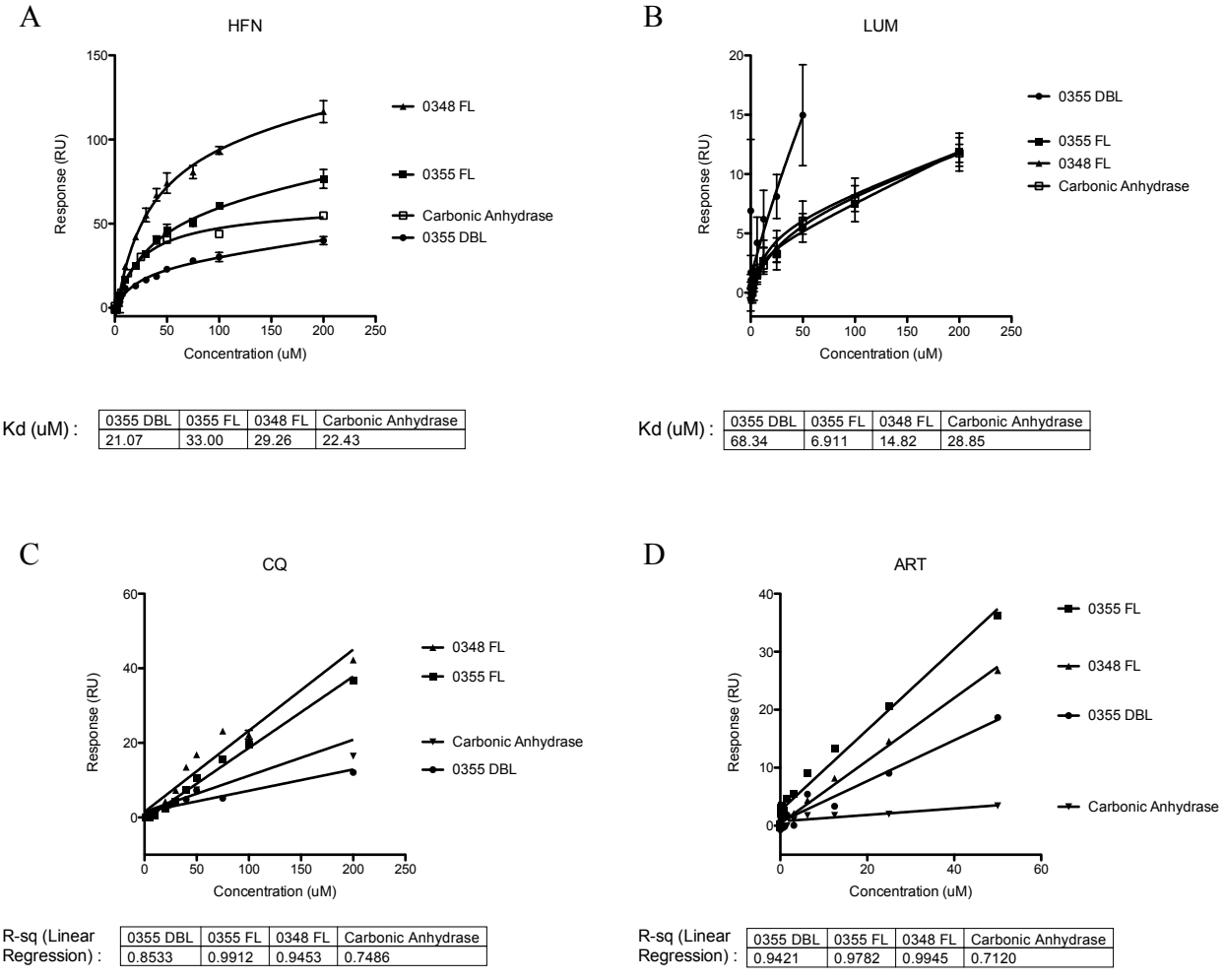


Figure B2. Drug binding to recombinant proteins by Biacore. PF10_0348 full length (0348 FL), PF10_0355 full length (0355 FL), PF10_0355 DBL domain (0355 DBL) and Carbonic Anhydrase proteins were analyzed for their binding to various antimalarials. Binding was measured between each protein and A) halofantrine (HFN), B) lumefantrine (LUM), C) chloroquine (CQ), and D) artemisinin (ART). Response units (RU) are plotted versus drug concentration. K_d values were calculated for binding of HFN and LUM; because CQ and ART showed a linear relationship between RU and drug concentration, K_d values were not calculated and R-squared values for linear fit are shown instead.



Kd (uM) :	0355 DBL	0355 FL	0348 FL	Carbonic Anhydrase
	21.07	33.00	29.26	22.43

Kd (uM) :	0355 DBL	0355 FL	0348 FL	Carbonic Anhydrase
	68.34	6.911	14.82	28.85

R-sq (Linear Regression) :	0355 DBL	0355 FL	0348 FL	Carbonic Anhydrase
	0.8533	0.9912	0.9453	0.7486

R-sq (Linear Regression) :	0355 DBL	0355 FL	0348 FL	Carbonic Anhydrase
	0.9421	0.9782	0.9945	0.7120

References

1. Shevchenko, A.*, et al.* (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry* 68, 850-858

2. Peng, J. and Gygi, S.P. (2001) Proteomics: the move to mixtures. *Journal of mass spectrometry : JMS* 36, 1083-1091

3. Jimmy K. Eng, A.L.M., John R. Yates III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* 5, 976-989

4. Johnsson, B.*, et al.* (1991) Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Analytical biochemistry* 198, 268-277

APPENDIX C

Supplemental Figures for Chapter Four:

Van Tyne et al. (in prep). Monitoring antimalarial drug response in *Plasmodium falciparum* field isolates using an *ex vivo* DAPI assay

Figure C1. Artemisinin IC_{50}s in Senegalese parasites tested with the DAPI *ex vivo* drug assay in 2008, 2009 and 2010. Scatter plots show log_{10}-transformed IC_{50}s and horizontal lines indicate median IC_{50} for each year. Median IC_{50}s were 3.4nM in 2008, 8.6nM in 2009 and 9.5nM in 2010.

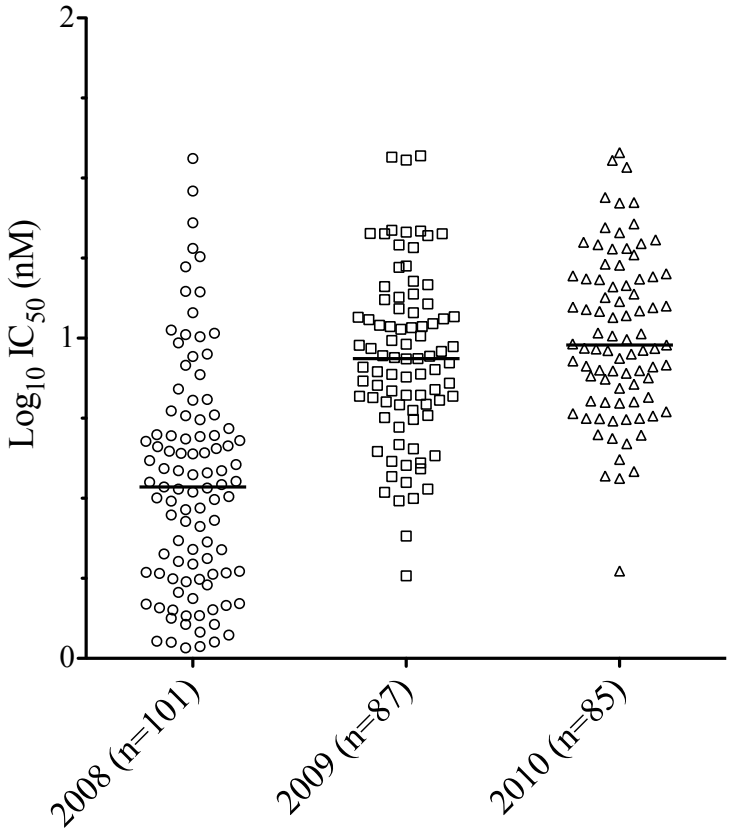


Figure C2. Comparison of *ex vivo* with *in vitro* IC_{50}s for artemisinin among 24 culture-adapted parasites collected in 2009. All parasites were derived from monoclonal infections. *Ex vivo* IC_{50}s were measured using the DAPI *ex vivo* assay and *in vitro* IC_{50}s were measured by radiolabeled hypoxanthine incorporation. Mean *in vitro* IC_{50}s are plotted with error bars showing the standard error of at least two biological replicates. ρ denotes the Pearson correlation coefficient.